METHODS FOR TREATING PATIENTS AND IDENTIFYING THERAPETICS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of U.S. Patent Application No. 10/274,177, filed October 18, 2002, which is a continuation-in-part of U.S. Patent Application No. 10/229,345, filed August 26, 2002, and both of the aforementioned patent applications are incorporated herein by reference. This application claims the benefit of the filing date of U.S. Provisional Patent Application No. 60/406,296, filed August 27, 2002, and incorporated herein by reference.

FUNDING

Work described herein was funded, in part, by grant number 1 U01 CA-88130-01 from the National Cancer Institute. The United States government has certain rights in the invention.

BACKGROUND

Colorectal cancer, also referred to herein as colon cancer, is the second leading cause of cancer mortality in the adult American population. An estimated 135,000 new cases of colon cancer occur each year. Although many people die of colon cancer, early stage colon cancers are often treatable by surgical removal (resection) of the affected tissue. Surgical treatment can be combined with chemotherapeutic agents to achieve an even higher survival rate in certain colon cancers. However, the survival rate drops to 5% or less over five years in patients with metastatic (late stage) colon cancer.

Effective screening and early identification of affected patients coupled with appropriate therapeutic intervention is proven to reduce the number of colon cancer mortalities. It is estimated that 74,000,000 older Americans would benefit from regular screening for colon cancer and precancerous colon adenomas (together, adenomas and colon cancers may be referred to as colon neoplasias). However, present systems for screening for colon neoplasia are inadequate. For example, the Fecal Occult Blood Test involves testing a stool sample from a patient for the presence of blood. This test is relatively simple and inexpensive, but it often fails to detect colon neoplasia (low sensitivity) and often even when blood is detected in the stool, a colon neoplasia is not

present (low specificity). Flexible sigmoidoscopy involves the insertion of a short scope into the rectum to visually inspect the lower third of the colon. Because the sigmoidoscope is relatively short, it is also a relatively uncomplicated diagnostic method. However, nearly half of all colon neoplasia occurs in the upper portions of the colon that can not be viewed with the sigmoidoscope. Colonoscopy, in which a scope is threaded through the entire length of the colon, provides a very reliable method of detecting colon neoplasia in a subject, but colonoscopy is costly, time consuming and requires sedation of the patient.

Modern molecular biology has made it possible to identify proteins and nucleic acids that are specifically associated with certain physiological states. These molecular markers have revolutionized diagnostics for a variety of health conditions ranging from pregnancy to viral infections, such as HIV.

Researchers generally identify molecular markers for a health condition by searching for genes and proteins that are expressed at different levels in one health condition versus another (e.g. in pregnant women versus women who are not pregnant). Traditional methods for pursuing this research, such as Northern blots and reverse transcriptase polymerase chain reaction, allow a researcher to study only a handful of potential molecular markers at a time. Microarrays, consisting of an ordered array of hundreds or thousands of probes for detection of hundreds or thousands of gene transcripts, allow researchers to gather data on many potential molecular markers in a single experiment. Researchers now face the challenge of sifting through large quantities of microarray-generated gene expression data to identify genes that may be of genuine use as molecular markers to distinguish different health conditions.

Improved systems for identifying high quality candidate molecular markers in large volumes of gene expression data may help to unlock the power of such tools and increase the likelihood of identifying a molecular marker for important disease states, such as colon neoplasia. Effective molecular markers for colon neoplasia could potentially revolutionize the diagnosis, management and overall health impact of colon cancer. In addition, molecular markers may be used in screening for, generating and targeting therapeutic agents for colon cancer.

BRIEF SUMMARY

This application is based at least in part on the selection of useful molecular targets for therapeutic intervention in treating neoplasia. Colon neoplasia is a multi-stage process involving progression from normal healthy tissues to the development of pre-cancerous colon adenomas to more invasive stages of colon cancer such as the Dukes A and Dukes B stages and finally to metastatic stages such as Dukes C and Dukes D stages of colon cancer.

In one aspect, this application provides molecular markers that are useful in the detection or diagnosis of colon neoplasia. In certain embodiments, molecular markers described in the application are helpful in distinguishing normal subjects from those who are likely to develop colon neoplasia or are likely to harbor a colon adenoma. In other aspects the invention provides molecular markers that may be useful in distinguishing subjects who are either normal or precancerous from those who have colon cancer. In another embodiment, the application provides markers that help in staging the colon cancer in patients. In still other embodiments the application contemplates the use of one or more of the molecular markers described herein for the detection, diagnosis, and staging of colon neoplasias. In certain embodiments, one or more markers for colon neoplasia disclosed herein may be used for identifying or targeting antineoplastic agents directed against colon neoplasia.

In certain aspects the application provides methods for inhibiting the growth or proliferation of a colon neoplasia in a subject, the method comprising administering to the subject an agent that decreases the amount of a polypeptide present in or produced by the colon neoplasia, said polypeptide selected from among: ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7 and ColoUp8. Optionally, the polypeptide is a secreted polypeptide, such as certain ColoUp1 or ColoUp2 polypeptides. Optionally, the polypeptide is a transmembrane polypeptide, such as certain ColoUp3 polypeptides. Optionally, the polypeptide is an intracellular polypeptide, such as ColoUp4, ColoUp5 or ColoUp6. Optionally, the agent is an siRNA probe that hybridizes to an mRNA encoding a polypeptide selected from among: ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7 and ColoUp8. In preferred embodiments, the siRNA probe hybridizes to a nucleic acid that is at least 90%, 95%,

98%, 99% or 100% identical to a nucleic acid sequence of one of SEQ ID Nos. 4, 5 and 7-12. Optionally, the agent is an antisense probe that hybridizes to a nucleic acid encoding a polypeptide selected from among: ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7 and ColoUp8. In preferred embodiments, the antisense probe hybridizes to a nucleic acid that is at least 90%, 95%, 98%, 99% or 100% identical to a nucleic acid sequence of one of SEQ ID Nos. 4, 5 and 7-12. In certain embodiments, the agent comprises a nucleic acid vector that causes the production of a siRNA or an antisense probe that hybridizes to a nucleic acid encoding a polypeptide selected from among: ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7 and ColoUp8.

In certain aspects, the application provides a method for inhibiting the growth or proliferation of a cell of a colon neoplasia in a subject, the method comprising administering to the subject an agent that binds to and antagonizes a polypeptide selected from among: ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7 and ColoUp8. In some embodiments, the agent comprises an antibody that binds to a polypeptide selected from among ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7 and ColoUp8. Optionally, the antibody binds to a polypeptide selected from among SEQ ID Nos. 1-3, 13, 14 and 16-21. Optionally, the antibody is a monoclonal antibody, a polyclonal antibody or a single chain antibody. Optionally, the antibody is a humanized antibody. In certain embodiments, the agent is a small molecule that binds to a polypeptide selected from among: SEQ ID Nos. 1-3, 13, 14 and 16-21, and preferably a small molecule that inhibits an activity of a polypeptide selected from among SEQ ID Nos. 1-3, 13, 14 and 16-21. For example, an agent may inhibit receptor binding (which may be assayed as cell surface binding) by a secreted polypeptide (e.g., SEQ ID Nos. 1, 2, 3 and 21). An agent may inhibit cadherin binding or intracellular signaling by ColoUp3. An agent may inhibit DNA binding and/or multimerization by ColoUp4 and ColoUp5. An agent may inhibit cytokeratin filament formation by ColoUp6.

In certain aspects, molecular markers of colon neoplasia may be used to target therapeutic agents to cells of a colon neoplasia. In certain embodiments, a therapeutic agent that is targeted to a colon neoplasia comprises a targeting moiety and an active

moiety, wherein the targeting moiety binds to a polypeptide selected from among ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7 and ColoUp8 and wherein the active moiety facilitates the killing or growth inhibition of a cell of a colon neoplasia. Optionally, the targeting moiety comprises an antibody. In preferred embodiments, the antibody binds to a polypeptide selected from among SEQ ID Nos. 1-3, 13, 14 and 16-21. Optionally, the antibody is selected from among: a monoclonal antibody, a polyclonal antibody, a single chain antibody. In certain embodiments, the antibody is a humanized antibody. The active moiety may be, for example, a toxin, a chemotherapeutic agent, or an agent that sensitizes the cell to a chemotherapeutic agent or radiation. In a preferred embodiment, the targeting moiety binds to a protein that is associated with the cell surface, and particularly ColoUp3, however, secreted markers may also be used, as such markers may have high local concentrations within the neoplasia and may adhere to the extracellular matrix in the neoplasia. Intracellular markers may also have high local concentrations in the neoplasia as a result of cell lysis. In addition, a therapeutic agent may comprise a moiety for intracellular targeting, such as an HIV tat protein, a porin, etc.

In certain embodiments, the application provides methods of identifying a candidate agent for treating colon cancer, the method comprising: identifying a candidate agent that binds to and/or inhibits an activity of a polypeptide selected from among: ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7 and ColoUp8. In certain embodiments, the method may further comprise testing the candidate agent for antineoplastic effects on a cell of a colon neoplasia or a cell of a cell line derived from a colon neoplasia. The method may further comprise testing the candidate agent for antineoplastic effects on a mouse xenograft comprising cells of a human colon cancer or cells of a cell line derived from a colon cancer cell line. The candidate agent may be essentially any molecule or complex material of interest, including, for example, a siRNA probe, an antisense probe, an antibody and a small molecule.

In one aspect the application provides a method of screening a subject for a condition associated with increased levels of one or more molecular markers that are indicative of colon neoplasia such as for example ColoUp1-ColoUp8 and osteopontin. In

a preferred embodiment, the application provides a method for screening a subject for conditions associated with secreted markers such as ColoUp1 or ColoUp2, by detecting in a biological sample an amount of ColoUp1 or ColoUp2 and comparing the amount of ColoUp1 and ColoUp2 found in the subject to one or more of the following: a predetermined standard, the amount of ColoUp1 or ColoUp2 detected in a normal sample from the subject, the subject's historical baseline level of ColoUp1 or ColoUp2, or the ColoUp1 or ColoUp2 level detected in a different, normal subject (a control subject). Detection of a level of ColoUp1 and ColoUp2 in the subject that is greater than that of the predetermined standard or that is increased from a subject's past baseline is indicative of a condition such as colon neoplasia. In certain aspects, an increase in the amount of ColoUp1 or ColoUp2 as compared to the subject's historical baseline would be indicative of a new neoplasia, or progression of an existing neoplasia. Similarly, a decrease in the amount of ColoUp1 or ColoUp2 as compared to the subject's historical baseline would be indicative of regression on an existing neoplasia

In one aspect the molecular markers described herein are encoded by a nucleic acid sequence that is at least 90%, 95%, 98%, 99%, 99.3%, 99.5% or 99.7% identical to the nucleic acid sequence of SEQ ID Nos: 4-12, and more preferably to the nucleic acid sequences as set forth in SEQ ID Nos: 4-5. In another aspect, the application provides markers that are encoded by a nucleic acid sequence that hybridizes under high stringency conditions to the nucleic acid sequences of SEQ ID Nos: 4-12, more preferably to the nucleic acid sequences as set forth in SEQ ID Nos: 4-5.

In another aspect the application provides molecular markers that are diagnostic of colon neoplasia, said markers having an amino acid sequence that is at least 90%, 95%, 98%, 99%, 99.3%, 99.5% or 99.7% identical to the amino acid sequence as set forth in SEQ ID Nos: 1-3 or 13-20, more preferably the amino acid sequence as set forth in SEQ ID Nos: 3 and 14.

In one aspect, the application provides methods for detecting secreted polypeptide forms of a ColoUp1 – ColoUp8 polypeptide or osteopontin in biological samples. In other aspects, the application provides methods for imaging a colon neoplasia by targeting antibodies to any one of the markers ColoUp1 through ColoUp8 described herein, and in preferred embodiments, the antibodies are targeted to ColoUp3. In certain

aspects, the application provides methods for administering a imaging agent comprising a targeting moiety and an active moiety. The targeting moiety may be an antibody, Fab, F(Ab)2, a single chain antibody or other binding agent that interacts with an epitope specified by a polypeptide sequence having an amino acid sequence as set forth in SEQ ID Nos: 1-3 and 13-20. The active moiety may be a radioactive agent, such as radioactive technetium, radioactive indium, or radioactive iodine. The imaging agent is administered in an amount effective for diagnostic use in a mammal such as a human and the localization and accumulation of the imaging agent is then detected. The localization and accumulation of the imaging agent may be detected by radioscintigraphy, nuclear magnetic resonance imaging, computed tomography or positron emission tomography.

In a preferred embodiment, the application provides methods for detecting a polypeptide comprising an amino acid sequence as set forth in one of SEQ ID Nos: 1-3. As will be apparent to the skilled artisan, the molecular markers described herein may be detected in a number of ways such as by various assays, including antibody-based assays. Examples of antibody-based assays include immunoprecipitation assays, Western blots, radioimmunoassays or enzyme-linked immunosorbent assays (ELISAs). Molecular markers described herein may be detected by assays that do not employ an antibody, such as by methods employing two-dimensional gel electrophoresis, methods employing mass spectroscopy, methods employing suitable enzymatic activity assays, etc. In a preferred embodiment the application provides methods for the detection of secreted markers such as ColoUp1 or ColoUp2 polypeptides in blood, blood fractions (such as blood serum or blood plasma), urine or stool samples. Increased levels of these markers may be associated with a number of conditions such as for example colon neoplasia, including colon adenomas, colon cancer, and metastatic colon cancer. In certain aspects the application provides methods including the detection of more than one marker that is indicative of colon neoplasia such as methods for detecting both ColoUp1 and ColoUp2. In yet another aspect, combinations of the ColoUp markers may be useful, for instance, a combination of tests including testing biological samples for secreted markers such as ColoUp1 or ColoUp2 in combination with testing for transmembrane markers such as ColoUp3 as targets for imaging agents..

In yet another aspect, the application provides a method of determining whether a subject is likely to develop colon cancer or is more likely to harbor a precancerous colon adenoma by detecting the presence or absence of the molecular markers as set forth in SEQ ID Nos: 1-3. Detection of combinations of these markers is also helpful in staging the colon neoplasias.

In yet another aspect, the application provides markers that are useful in distinguishing normal and precancerous subjects from those subjects having colon cancer. In certain embodiments, the application contemplates determining the levels of markers provided herein such as ColoUp1 through ColoUp8 and osteopontin. In one aspect, markers such as ColoUp6 and osteopontin are helpful in distinguishing between the category of patients that are normal or have precancerous colon adenomas and the category of patients having colon cancer. In another aspect, the application provides detection of one or more of said markers in determining the stages of colon neoplasia.

In certain aspect, the invention provides an immunoassay for determining the presence of any one of the polypeptides having an amino acid sequence as set forth in SEQ ID Nos: 1-3 and 13-20, more preferably any one of the polypeptides having an amino acid sequence as set forth in SEQ ID Nos: 1-3 in a biological sample. The method includes obtaining a biological sample and contacting the sample with an antibody specific for a polypeptide having an amino acid sequence as set forth in SEQ ID Nos: 1-3 and detecting the binding of the antibody.

In some aspects, the application provides methods for the detection of a molecular marker in a biological sample such as blood, including blood fractions such as serum or plasma. For instance, the blood sample obtained from a patient may be further processed such as by fractionation to obtain blood serum, and the serum may then be enriched for certain polypeptides. The serum so enriched is then contacted with an antibody that is reactive with an epitope of the desired marker polypeptide.

In yet another embodiment, the application provides methods for determining the appropriate therapeutic protocol for a subject. For example detection of a colon neoplasia provides the treating physician valuable information in determining whether intensive or invasive protocols such as colonoscopy, surgery or chemotherapy would be needed for effective diagnosis or treatment. Such detection would be helpful not only for patients

not previously diagnosed with colon neoplasia but also in those cases where a patient has previously received or is currently receiving therapy for colon cancer, the presence or absence or a change in the level of the molecular markers set forth herein may be indicative that the subject is likely to have a relapse or a progressive, or a persistent colon cancer.

In certain aspects, the application provides molecular markers of colon neoplasia such as ColoUp1 through ColoUp8. In certain instances these markers are secreted proteins such as ColoUp1, ColoUp2 and osteopontin, and are useful for detecting and diagnosing colon neoplasia. In other aspects, these markers may be transmembrane proteins such as ColoUp3 and may be useful as targets for imaging agents, e.g. as targets to label cells of a neoplasia.

In one aspect, the application provides isolated, purified or recombinant polypeptides having an amino acid sequence that is at least 90%, 95% or 98-99% identical to an amino acid sequence as set forth in SEQ ID Nos: 1-3 or an amino acid sequence as set forth in SEQ ID Nos: 13-20. In a more preferred embodiment, the application provides an amino acid sequence that is at least 90%, 95%, 98-99%, 99.3%, 99.5% or 99.7% identical to the amino acid sequence as set forth in SEQ ID No: 3 or SEQ ID No: 14. The application also provides fusion proteins comprising the ColoUp proteins described herein fused to a heterologous protein. In certain embodiments, such polypeptides are useful, for example, for generating antibodies or for use in screening assays to identify candidate therapeutics.

In other aspects the application provides for nucleic acid sequences encoding the polypeptides as set forth in SEQ ID Nos: 1-3 and 13-20. In one aspect the application provides nucleic acids comprising nucleic acid sequences that are at least 90%, 95%, 98-99%, 99.3%, 99.5% or 99.7% identical to the nucleic acid sequence in SEQ ID Nos: 4-12, more preferably 4-5. Also contemplated herein are vectors comprising the nucleic acid sequences set forth in SEQ ID Nos: 4-12, more preferably SEQ ID Nos: 4-5, and host cells expressing the nucleic acid sequences.

In another aspect, the application provides an antibody that interacts with an epitope specified by one of SEQ ID Nos: 1-3 and 13-20 or portions thereof, more preferably SEQ ID Nos: 1-3 or portions thereof. In a preferred embodiment the antibody

is useful for detecting colon adenomas and interacts with an epitope specified by one of SEQ ID Nos: 1-3. In certain aspects the application provides for generating such antibodies, including methods for generating monoclonal and polyclonal antibodies, as well as methods for generating other types of antibodies. In other aspects, the application also provides a hybridoma cell line capable of producing an antibody that interacts with an epitope specified by SEQ ID Nos: 1-3 and 13-20, more preferably SEQ ID Nos: 1-3, or portions thereof In yet other embodiments, the antibody may be a single chain antibody.

In yet other embodiments, the application provides a kit for detecting colon neoplasia in a biological sample, Such kits include one or more antibodies that are capable of interacting with an epitope specified by one of SEQ ID Nos: 1-3 and 13-20, more preferably with an epitope specified by one of SEQ ID Nos: 1-3. In more preferred embodiments, the antibodies may be detectably labeled, such as for example with an enzyme, a fluorescent substance, a chemiluminescent substance, a chromophore, a radioactive isotope or a complexing agent.

In certain embodiments, the application provides the identity of ColoUp1 and ColoUp2 polypeptides that are secreted into the serum in vivo, and that are secreted across the apical and basolateral cell surfaces in cultured intestinal cells. Accordingly, in certain embodiments, the application provides methods for detecting whether a subject to likely to have a colon neoplasia comprising: a) obtaining a biological sample from said subject; and b) detecting one or more polypeptides selected from among: one or more secreted ColoUp1 polypeptides and one or more secreted ColoUp2 polypeptides, wherein the presence of said one or more polypeptides is indicative of colon neoplasia.

In certain embodiments, a secreted ColoUp2 polypeptide is selected from among: a) a secreted polypeptide produced by the expression of a nucleic acid that is at least 95% identical to the amino acid sequence of SEQ ID No: 5; b) a secreted polypeptide produced by the expression of a nucleic acid that is a naturally occurring variant of SEQ ID No: 5; c) a secreted polypeptide produced by the expression of a nucleic acid that hybridizes under stringent conditions to a nucleic acid sequence of SEQ ID No: 5; d) a secreted polypeptide having a sequence that is at least 95% identical to the amino acid sequence of SEQ ID No: 3; and e) a secreted polypeptide having a sequence that is at

least 95% identical to the amino acid sequence of SEQ ID No: 21. Optionally, the secreted ColoUp2 polypeptide is produced by the expression of a nucleic acid having the sequence of SEQ ID No: 5, and preferably the secreted ColoUp2 polypeptide is produced by the expression of a nucleic acid sequence that is at least 98%, 99% or 100% identical to the nucleic acid sequence of SEQ ID No: 5. In certain embodiments, the secreted ColoUp2 polypeptide has an amino acid sequence that is at least 98%, 99% or 100% identical to an amino acid sequence selected from among SEQ ID No: 3 and SEQ ID No:21. In certain embodiments, the secreted ColoUp1 polypeptide is selected from among: a) a secreted polypeptide produced by the expression of a nucleic acid that is at least 95% identical to the amino acid sequence of SEQ ID No: 4; b) a secreted polypeptide produced by the expression of a nucleic acid that is a naturally occurring variant of SEQ ID No: 4; c) a secreted polypeptide produced by the expression of a nucleic acid that hybridizes under stringent conditions to a nucleic acid sequence of SEQ ID No: 4; d) a secreted polypeptide having a sequence that is at least 95% identical to the amino acid sequence of SEQ ID No: 1; and e) a secreted polypeptide having a sequence that is at least 95% identical to the amino acid sequence of SEQ ID No: 2. Optionally, the secreted ColoUp1 polypeptide is produced by the expression of a nucleic acid having a sequence that is at least 95%, 98, 99% or 100% identical to the nucleic acid sequence of SEQ ID No: 4. Preferably, the secreted ColoUp1 polypeptide has an amino acid sequence that is at least 95%, 98%, 99% or 100% identical to an amino acid sequence selected from among SEQ ID No: 1 and SEQ ID No:2. Optionally, for detection of basolaterally secreted ColoUp1 or ColoUp2 polypeptides, the biological sample is a blood sample or a fraction derived from blood, such as serum, plasma, cells, or a fraction enriched for apically secreted ColoUp1 or ColoUp2 polypeptide. Optionally, for detection of basolaterally secreted ColoUp1 or ColoUp2 polypeptides, the biological sample is a urine sample or a fraction derived from urine. Optionally, for detection of apically secreted ColoUp1 or ColoUp2 polypeptides, the biological sample is derived from the inner wall and/or lumen of the intestinal tract, such as intestinal mucous or other fluid, excreted stool and stool removed from within the colon. In certain embodiments, the polypeptide is detected by an assay that employs an antibody, such as an immunoprecipitation assay, a Western blot, a radioimmunoassays or an enzyme-linked

immunosorbent assay (ELISA). Optionally, an assay comprises contacting the biological sample with an antibody that interacts with a secreted ColoUp1 polypeptide or a secreted ColoUp2 polypeptide. An antibody may, for example, interact with an epitope of an amino acid sequence selected from among: SEQ ID No: 1 and SEQ ID No: 2. An antibody may, for example, interact with an epitope of an amino acid sequence selected from among: SEQ ID No: 3 and SEQ ID No: 21. Optionally, the antibody is detectably labeled, such as with an enzyme, a fluorescent substance, a chemiluminescent substance, a chromophore, a radioactive isotope or a complexing agent. Optionally, the amount of at least one secreted ColoUp1 polypeptide and/or at least one secreted ColoUp2 polypeptide in the biological sample is compared to a predetermined standard (e.g., a known amount of purified ColoUp1 or ColoUp2 polypeptide). Optionally, the amount of at least one secreted ColoUp1 polypeptide and/or at least one secreted ColoUp2 polypeptide in the biological sample is compared to the subject's historical baseline. In certain embodiments, the presence of at least one secreted ColoUp1 polypeptide and/or at least one secreted ColoUp2 polypeptide is indicative that the subject is likely to harbor a colon adenoma or a colon cancer. In certain embodiments, the presence of at least one secreted ColoUp1 polypeptide and/or at least one secreted ColoUp2 polypeptide may be used in determining the therapeutic protocol to be administered to a subject having a colon neoplasia, and the subject may not have been previously diagnosed with colon cancer or the subject may have previously received or is currently receiving a therapy for colon cancer, wherein the presence of at least one secreted ColoUp1 polypeptide and/or at least one secreted ColoUp2 polypeptide indicates that the subject is likely to have a relapse or a persistent or progressive colon cancer. The detection of said secreted polypeptide may indicate the presence of a variety of neoplasias in a subject, such as a colon adenoma, a colon cancer and a metastatic colon cancer. Optionally, a method involves detecting both at least one secreted ColoUp1 polypeptide and at least one secreted ColoUp2 polypeptide in the biological sample.

In certain embodiments, the application provides kits for detecting one or more molecular markers of colon neoplasia in a biological sample. A kit may comprise a) an antibody which interacts with an epitope of a secreted ColoUp1 polypeptide or a secreted ColoUp2 polypeptide; and b)instructions for use. Optionally, the antibody interacts with

an epitope of a polypeptide selected from among: the polypeptide of SEQ ID No:1, the polypeptide of SEQ ID No:2, the polypeptide of SEQ ID No:3 and the polypeptide of SEQ ID No:21. Optionally, the antibody is detectably labeled.

In certain embodiments, the application provides a novel purified polypeptide, which is a portion of ColoUp2 that is found in serum. Such a polypeptide may consist essentially of an amino acid sequence that is at least 95%, 98%, 99% or 100% identical to the sequence of SEQ ID No: 21. By "consisting essentially" is meant that there may be, in addition to the indicated amino acid sequence, a variety of modifications, such as phosphorylations, glycosylations, disulfide bonds, unusual or modified amino acids, etc.

In certain embodiments, the application provides novel fusion proteins comprising a first polypeptide domain and a second polypeptide domain, wherein the first polypeptide domain consists essentially of an amino acid sequence that is at least 95%, 98%, 99% or 100% identical to an amino acid sequence of SEQ ID No. 21. The second polypeptide domain may be a domain selected from the group consisting of: a detection domain, a purification domain and an antigenic domain.

In certain embodiments, the application provides antibodies that bind specifically to a ColoUp2 polypeptide consisting essentially of the amino acid sequence of SEQ ID No: 21. The antibody may binds the ColoUp2 polypeptide with a dissociation constant of less than 10^{-6}M, 10^{-7}M, 10^{-8}M or 10^{-9}M. The antibody may be essentially any type of antibody, including polyclonal, monoclonal, and single chain antibodies, or other fragments. For diagnostic use, there may be little benefit to having a humanized antibody, however, humanized antibodies are highly desirable for therapeutic uses. Preferably, a diagnostic antibody is effective for detecting the ColoUp2 polypeptide in a biological sample, such as a blood, stool or urine sample, or a fraction thereof. Optionally, the antibody is effective for detecting the ColoUp2 polypeptide in a sample comprising cells from a colon neoplasia. The application further provides methods for making such antibodies in a variety of ways. For example, a monoclonal antibody may be produced in a method comprising: (a) administering to a mouse an amount of an immunogenic composition comprising the ColoUp2 polypeptide effective to stimulate a detectable immune response; (b) obtaining antibody-producing cells from the mouse and fusing the antibody–producing cells with myeloma cells to obtain antibody-producing

hybridomas; (c) testing the antibody-producing hybridomas to identify a preferred hybridoma, wherein the preferred hybridoma is a hybridoma that produces a monocolonal antibody that binds specifically to the ColoUp2 polypeptide; (d) culturing the preferred hybridoma cell culture that produces the monoclonal antibody that binds specifically to the ColoUp2 polypeptide; and (e) obtaining the monoclonal antibody that binds specifically to the ColoUp2 polypeptide from the cell culture. Optionally, the antibody-producing hybridomas comprises testing whether the antibody-producing hybridomas produce an antibody that binds to the ColoUp2 polypeptide in an assay selected from the group consisting of: an enzyme-linked immunosorbent assay, a Bia-core assay and an immunoprecipitation assay.

The embodiments and practices of the present invention, other embodiments, and their features and characteristics, will be apparent from the description, figures and claims that follow, with all of the claims hereby being incorporated by this reference into this Summary.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the amino acid sequences (SEQ ID NOs: 1 and 2) of secreted ColoUp1 protein. **A.** An N-terminal signal peptide is cleaved between amino acids 30-31 of the full-length ColoUp1 protein; **B.** An N-terminal signal peptide is cleaved between amino acids 33-34 of the full-length ColoUp1protein.

Figure 2 shows the amino acid sequence (SEQ ID NO: 3) of secreted ColoUp2 protein.

Figure 3 shows the nucleic acid sequence (SEQ ID NO: 4) of ColoUp1.

Figure 4 shows the nucleic acid sequence (SEQ ID NO: 5) of ColoUp2.

Figure 5 shows the nucleic acid sequence (SEQ ID NO: 6) of Osteopontin.

Figure 6 shows the nucleic acid sequence (SEQ ID NO: 7) of ColoUp3.

Figure 7 shows the nucleic acid sequence (SEQ ID NO: 8) of ColoUp4.

Figure 8 shows the nucleic acid sequence (SEQ ID NO: 9) of ColoUp5.

Figure 9 shows the nucleic acid sequence (SEQ ID NO: 10) of ColoUp6.

Figure 10 shows the nucleic acid sequence (SEQ ID NO: 11) of ColoUp7.

Figure 11 shows the nucleic acid sequence (SEQ ID NO: 12) of ColoUp8.

Figure 12 shows the amino acid sequence (SEQ ID NO: 13) of full-length ColoUp1 protein.

Figure 13 shows the amino acid sequence (SEQ ID NO: 14) of full-length ColoUp2 protein.

Figure 14 shows the amino acid sequence (SEQ ID NO: 15) of full-length Osteopontin protein.

Figure 15 shows the amino acid sequence (SEQ ID NO: 16) of full-length ColoUp3 protein.

Figure 16 shows the amino acid sequence (SEQ ID NO: 17) of full-length ColoUp4 protein.

Figure 17 shows the amino acid sequence (SEQ ID NO: 18) of full-length ColoUp5 protein.

Figure 18 shows the amino acid sequence (SEQ ID NO: 19) of full-length ColoUp6 protein.

Figure 19 shows the amino acid sequence (SEQ ID NO: 20) of full-length ColoUp8 protein.

Figure 20 is a graphical display of ColoUp1 expression levels measured by micro-array profiling in different samples. **A.** In normal colon epithelial strips, normal liver, and colonic muscle; **B.** In premalignant colon adenomas as well as in colon cancers of Dukes stages B, Dukes stage C, and Duke stages D; **C.** In colon cancer liver metastasis; **D**. In colon cancer cell lines, colon cancer xenografts grown in athymic mice, MSI cell lines, and V330 cell lines treated with TGFβ.

Figure 21 is a graphical display of ColoUp2 expression levels measured by micro-array profiling in different samples. **A.** In normal colon epithelial strips, normal liver, and colonic muscle; **B.** In premalignant colon adenomas as well as in colon cancers of Dukes stages B, Dukes stage C, and Duke stages D; **C.** In colon cancer liver metastasis; **D**. In colon cancer cell lines, colon cancer xenografts grown in athymic mice, MSI cell lines, and V330 cell lines treated with TGFβ.

Figure 22 is a graphical display of Osteopontin expression levels measured by micro-array profiling in different samples. **A.** In normal colon epithelial strips, normal liver, and colonic muscle; **B.** In premalignant colon adenomas as well as in colon cancers of Dukes stages B, Dukes stage C, and Duke stages D; **C.** In colon cancer liver metastasis; **D**. In colon cancer cell lines, colon cancer xenografts grown in athymic mice, MSI cell lines, and V330 cell lines treated with TGFβ.

Figure 23 is a graphical display of ColoUp3 expression levels measured by micro-array profiling in different samples. **A.** In normal colon epithelial strips, normal liver, and colonic muscle; **B.** In premalignant colon adenomas as well as in colon cancers of Dukes stages B, Dukes stage C, and Duke stages D; **C.** In colon cancer liver metastasis; **D**. In colon cancer cell lines, colon cancer xenografts grown in athymic mice, MSI cell lines, and V330 cell lines treated with TGFβ.

Figure 24 is a graphical display of ColoUp4 expression levels measured by micro-array profiling in different samples. **A.** In normal colon epithelial strips, normal liver, and colonic muscle; **B.** In premalignant colon adenomas as well as in colon cancers of Dukes stages B, Dukes stage C, and Duke stages D; **C.** In colon cancer liver metastasis; **D.** In colon cancer cell lines, colon cancer xenografts grown in athymic mice, MSI cell lines, and V330 cell lines treated with TGFβ.

Figure 25 is a graphical display of ColoUp5 expression levels measured by micro-array profiling in different samples. **A.** In normal colon epithelial strips, normal liver, and colonic muscle; **B.** In premalignant colon adenomas as well as in colon cancers of Dukes stages B, Dukes stage C, and Duke stages D; **C.** In colon cancer liver metastasis; **D.** In colon cancer cell lines, colon cancer xenografts grown in athymic mice, MSI cell lines, and V330 cell lines treated with TGFβ.

Figure 26 is a graphical display of ColoUp6 expression levels measured by micro-array profiling in different samples. **A.** In normal colon epithelial strips, normal liver, and colonic muscle; **B.** In premalignant colon adenomas as well as in colon cancers of Dukes stages B, Dukes stage C, and Duke stages D; **C.** In colon cancer liver metastasis; **D.** In colon cancer cell lines, colon cancer xenografts grown in athymic mice, MSI cell lines, and V330 cell lines treated with TGFβ.

Figure 27 is a graphical display of ColoUp7 expression levels measured by micro-array profiling in different samples. **A.** In normal colon epithelial strips, normal liver, and colonic muscle; **B.** In premalignant colon adenomas as well as in colon cancers of Dukes stages B, Dukes stage C, and Duke stages D; **C.** In colon cancer liver metastasis; **D.** In colon cancer cell lines, colon cancer xenografts grown in athymic mice, MSI cell lines, and V330 cell lines treated with TGFβ.

Figure 28 is a graphical display of ColoUp8 expression levels measured by micro-array profiling in different samples. **A.** In normal colon epithelial strips, normal liver, and colonic muscle; **B.** In premalignant colon adenomas as well as in colon cancers of Dukes

stages B, Dukes stage C, and Duke stages D; **C.** In colon cancer liver metastasis; **D.** In colon cancer cell lines, colon cancer xenografts grown in athymic mice, MSI cell lines, and V330 cell lines treated with TGFβ.

Figure 29 shows northern blot analysis of ColoUp1 mRNA levels in normal colon tissues and colon cancer cell lines or tissues. **A.** In normal colon tissue samples and a group of colon cancer cell lines; **B.** and **C.** In normal colon tissues and colon neoplasias from 15 individuals with colon cancers and one individual with a colon adenoma.

Figure 30 shows detection of T7 epitope-tagged ColoUp1 protein levels in transfected FET cells and Vaco400 cells. **A.** Secretion of epitope-tagged ColoUp1 protein in V400 cell growth media by Western blot ("T" are transfectants with an epitope tagged ColoUp1 expression vector; "C" are transfectants with an empty control vector); **B.** Expression of T7 epitope-tagged ColoUp1 protein in transfected FET cells and V400 cells by Western blot (left panel), and secretion of epitope-tagged ColoUp1 protein in growth media by serial immunoprecipitation and Western blot (right panel)(Cell extract amounts loaded: FET = 75 mg/well; V400 = 31.1 mg/well; Volume of media used for immuno-precipitation = 1 ml of 20 ml).

Figure 31 shows northern blot analysis of ColoUp2 mRNA levels in normal colon tissue samples and a group of colon cancer cell lines (top panel). The bottom panel shows the ethidium bromide stained gel corresponding to the blot.

Figure 32 shows detection of V5 epitope-tagged ColoUp2 protein levels in transfected SW480 cells and Vaco400 cells (24 hours and 48 hours after trnasfection). Expression of epitope-tagged ColoUp2 protein in transfected cells by Western blot (right panel), and secretion of epitope-tagged ColoUp2 protein in growth media by serial immunoprecipitation and Western blot (left panel).

Figure 33 shows two northern blot analysis of ColoUp5 mRNA levels in normal colon tissues and a group of colon cancer cell lines (top panels). The bottom panels show the ethidium bromide stained gel corresponding to the blot.

Figure 34 illustrates an alignment of the human, mouse, and rat ColoUp5 (FoxQ1) amino acid sequences.

Figure 35 illustrates an alignment of the human, mouse, and rat ColoUp5 (FoxQ1) nucleic acid sequences.

Figure 36 shows a western blot of V5 tagged ColoUp2 protein detected by anti-V5 antibody. Lane 1: media supernate from SW480 colon cancer cells transfected with an empty expression vector. Lane 2: media supernate from ColoUp2-V5 expressing cells. Lane 3 : size markers. Lane 4 shows assay of serum from a mouse xenografted with control SW480 cells corresponding to lane 1. Lanes 5 and 6 show detection of circulating ColoUp2 proteins in blood from two mice bearing human colon cancer xenografts from ColoUp2-V5 expressing SW480 colon cells shown in lane 2. ColoUp2 is secreted as an 85KD and a companion 55KD size protein.

Figure 37 shows a western blot with anti-V5 antibody of V5 tagged ColoUp1 protein. Lane 1: media supernate from SW480 colon cancer cells transfected with an empty expression vector. Lane 2: media supernate from ColoUp1-V5 expressing SW480 cells. Lane 3 shows assay of serum from a mouse xenografted with control SW480 cells corresponding to lane 1. Lanes 4 shows detection of circulating ColoUp1 proteins in blood from a mouse bearing tumor xenografts from ColoUp1-V5 expressing SW480 cells shown in lane 2. Lane 5 : size markers.

Figure 38 shows, in the upper panel, the purification of ColoUp2 protein. Shown is a Coomassie blue staining of 250ng (lane 2a) and 500ng (lane 3a) of a purified ColoUp2 protein preparation. Size markers are in lane 1a. In the lower panel is shown a Coomassie blue stained gel showing purification of His-tagged ColoUp1 protein on Ni-

NTA beads. Lane 1:markers, Lane 2 media from mock transfected cells, Lane 3 purification of media from ColoUp1 transfected cells. Clearly shown is purification to homogeneity of the 180kd ColoUp protein.

Figure 39 shows, in the top panel, detection on an anti-V5 western of V5-tagged ColoUp2 protein. Lane 1: media from mock transfected Caco2 cells. Lane 2: detection of secreted ColoUp2 protein from transiently transfected Caco2 cells grown in standard culture dishes. Seen are the typical 85KD and 55KD secreted bands (the lane is heavily overloaded and minor degradation products are also visualized). Lane 3: molecular weight markers. Lanes 4-7: detection of ColoUp2 secreted into the basolateral compartment (lower chamber) of transiently transfected Caco2 grown as a monolayer on a transwell filter. Lanes 9-12 show the general absence of ColoUp2 in the corresponding apical apical compartment, with the exception of the 48 hour time point. The table shows the electrical resistance and transfection efficiency (gfp expression) measured at each time point. A dip in the electrical resistance at 48 hours suggests some leakiness of the monolayer at that time point.

Figure 40: Top panel shows detection on anti-V5 western of V5-tagged ColoUp1 protein. Control lane shows detection of purified recombinant ColoUp1. Identical bands are seen in media harvested on days 1-4 (lanes D1-D4) from both apical and basolateral compartments. The table shows the electrical resistance and transfection efficiency (gfp expression) measured at each time point.

Figure 41 shows the amino acid sequence of the approximately 55 kDa C-terminal fragment of ColoUp2 that is a prominent secreted and serum form of ColoUp2.

DETAILED DESCRIPTION

1. Definitions:

For convenience, certain terms employed in the specification, examples, and appended claims are collected here. Unless defined otherwise, all technical and scientific

terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

The terms "adenoma", "colon adenoma" and "polyp" are used herein to describe any precancerous neoplasia of the colon.

The term "antibody" as used herein is intended to include whole antibodies, e.g., of any isotype (IgG, IgA, IgM, IgE, etc), and includes fragments thereof which are also specifically reactive with a vertebrate, e.g., mammalian, protein. Antibodies can be fragmented using conventional techniques and the fragments screened for utility and/or interaction with a specific epitope of interest. Thus, the term includes segments of proteolytically-cleaved or recombinantly-prepared portions of an antibody molecule that are capable of selectively reacting with a certain protein. Non-limiting examples of such proteolytic and/or recombinant fragments include Fab, F(ab')2, Fab', Fv, and single chain antibodies (scFv) containing a V[L] and/or V[H] domain joined by a peptide linker. The scFv's may be covalently or non-covalently linked to form antibodies having two or more binding sites. The term antibody also includes polyclonal, monoclonal, or other purified preparations of antibodies and recombinant antibodies.

The term "colon" as used herein is intended to encompass the right colon (including the cecum), the transverse colon, the left colon and the rectum.

The terms "colorectal cancer" and "colon cancer" are used interchangeably herein to refer to any cancerous neoplasia of the colon (including the rectum, as defined above).

The term "ColoUpX" (e.g. ColoUp1, ColoUp2...ColoUp8) is used to refer to a nucleic acid encoding a ColoUp protein or a ColoUp protein itself, as well as distinguishable fragments of such nucleic acids and proteins, longer nucleic acids and polypeptides that comprise distinguishable fragments or full length nucleic acids or polypeptides, and variants thereof. Variants include polypeptides that are at least 90% identical to the relevant human ColoUp SEQ ID Nos. referred to in the application, and nucleic acids encoding such variant polypeptides. In addition, variants include different post-translational modifications, such as glycosylations, methylations, etc. Particularly

preferred variants include any naturally occurring variants, such as allelic differences, mutations that occur in a neoplasia and secreted or processed forms. The terms "variants" and "fragments" are overlapping.

As used herein, the phrase "gene expression" or "protein expression" includes any information pertaining to the amount of gene transcript or protein present in a sample, as well as information about the rate at which genes or proteins are produced or are accumulating or being degraded (eg. reporter gene data, data from nuclear runoff experiments, pulse-chase data etc.). Certain kinds of data might be viewed as relating to both gene and protein expression. For example, protein levels in a cell are reflective of the level of protein as well as the level of transcription, and such data is intended to be included by the phrase "gene or protein expression information". Such information may be given in the form of amounts per cell, amounts relative to a control gene or protein, in unitless measures, etc.; the term "information" is not to be limited to any particular means of representation and is intended to mean any representation that provides relevant information. The term "expression levels" refers to a quantity reflected in or derivable from the gene or protein expression data, whether the data is directed to gene transcript accumulation or protein accumulation or protein synthesis rates, etc.

The term "detection" is used herein to refer to any process of observing a marker, in a biological sample, whether or not the marker is actually detected. In other words, the act of probing a sample for a marker is a "detection" even if the marker is determined to be not present or below the level of sensitivity. Detection may be a quantitative, semi-quantitative or non-quantitative observation.

The terms "healthy", "normal" and "non-neoplastic" are used interchangeably herein to refer to a subject or particular cell or tissue that is devoid (at least to the limit of detection) of a disease condition, such as a neoplasia, that is associated with increased expression of a ColoUp gene. These terms are often used herein in reference to tissues and cells of the colon. Thus, for the purposes of this application, a patient with severe heart disease but lacking a ColoUp-associated disease would be termed "healthy".

The term "including" is used herein to mean, and is used interchangeably with, the phrase "including but not limited to".

As used herein, the term "nucleic acid" refers to polynucleotides such as deoxyribonucleic acid (DNA), and, where appropriate, ribonucleic acid (RNA). The term should also be understood to include analogs of either RNA or DNA made from nucleotide analogs, and, as applicable to the embodiment being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

The term "or" is used herein to mean, and is used interchangeably with, the term "and/or", unless context clearly indicates otherwise.

The term "percent identical" refers to sequence identity between two amino acid sequences or between two nucleotide sequences. Identity can each be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When an equivalent position in the compared sequences is occupied by the same base or amino acid, then the molecules are identical at that position; when the equivalent site occupied by the same or a similar amino acid residue (e.g., similar in steric and/or electronic nature), then the molecules can be referred to as homologous (similar) at that position. Expression as a percentage of homology/similarity or identity refers to a function of the number of identical or similar amino acids at positions shared by the compared sequences. Various alignment algorithms and/or programs may be used, including FASTA, BLAST or ENTREZ. FASTA and BLAST are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with, e.g., default settings. ENTREZ is available through the National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Md. In one embodiment, the percent identity of two sequences can be determined by the GCG program with a gap weight of 1, e.g., each amino acid gap is weighted as if it were a single amino acid or nucleotide mismatch between the two sequences.

The terms "polypeptide" and "protein" are used interchangeably herein.

The term "purified protein" refers to a preparation of a protein or proteins which are preferably isolated from, or otherwise substantially free of, other proteins normally associated with the protein(s) in a cell or cell lysate. The term "substantially free of other cellular proteins" (also referred to herein as "substantially free of other contaminating proteins") is defined as encompassing individual preparations of each of the component

proteins comprising less than 20% (by dry weight) contaminating protein, and preferably comprises less than 5% contaminating protein. Functional forms of each of the component proteins can be prepared as purified preparations by using a cloned gene as described in the attached examples. By "purified", it is meant, when referring to component protein preparations used to generate a reconstituted protein mixture, that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins (particularly other proteins which may substantially mask, diminish, confuse or alter the characteristics of the component proteins either as purified preparations or in their function in the subject reconstituted mixture). The term "purified" as used herein preferably means at least 80% by dry weight, more preferably in the range of 85% by weight, more preferably 95-99% by weight, and most preferably at least 99.8% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above.

A "recombinant nucleic acid" is any nucleic acid that has been placed adjacent to another nucleic acid by recombinant DNA techniques. A "recombinant nucleic acid" also includes any nucleic acid that has been placed next to a second nucleic acid by a laboratory genetic technique such as, for example, tranformation and integration, transposon hopping or viral insertion. In general, a recombined nucleic acid is not naturally located adjacent to the second nucleic acid.

The term "recombinant protein" refers to a protein that is produced by expression from a recombinant nucleic acid.

A "sample" includes any material that is obtained or prepared for detection of a molecular marker, or any material that is contacted with a detection reagent or detection device for the purpose of detecting a molecular marker.

A "subject" is any organism of interest, generally a mammalian subject, such as a mouse, and preferably a human subject.

2. Overview

In certain aspects, the invention relates to methods for determining whether a subject is likely or unlikely to have a colon neoplasia and markers that may be used to make such determination and to selected and/or target antineoplastic therapeutic agents. In other aspects, the invention relates to methods for determining whether a patient is likely or unlikely to have a colon cancer. In further aspects, the invention relates to methods for monitoring colon neoplasia in a subject. In further aspects, the invention relates to methods for staging a subject's colon neoplasia. A colon neoplasia is any cancerous or precancerous growth located in, or derived from, the colon. The colon is a portion of the intestinal tract that is roughly three feet in length, stretching from the end of the small intestine to the rectum. Viewed in cross section, the colon consists of four distinguishable layers arranged in concentric rings surrounding an interior space, termed the lumen, through which digested materials pass. In order, moving outward from the lumen, the layers are termed the mucosa, the submucosa, the muscularis propria and the subserosa. The mucosa includes the epithelial layer (cells adjacent to the lumen), the basement membrane, the lamina propria and the muscularis mucosae. In general, the "wall" of the colon is intended to refer to the submucosa and the layers outside of the submucosa. The "lining" is the mucosa.

Precancerous colon neoplasias are referred to as adenomas or adenomatous polyps. Adenomas are typically small mushroom-like or wart-like growths on the lining of the colon and do not invade into the wall of the colon. Adenomas may be visualized through a device such as a colonoscope or flexible sigmoidoscope. Several studies have shown that patients who undergo screening for and removal of adenomas have a decreased rate of mortality from colon cancer. For this and other reasons, it is generally accepted that adenomas are an obligate precursor for the vast majority of colon cancers.

When a colon neoplasia invades into the basement membrane of the colon, it is considered a colon cancer, as the term "colon cancer" is used herein. In describing colon cancers, this specification will generally follow the so-called "Dukes" colon cancer staging system. Other staging systems have been devised, and the particular system selected is, for the purposes of this disclosure, unimportant. The characteristics that the describe a cancer are of greater significance than the particular term used to describe a recognizable stage. The most widely used staging systems generally use at least one of

the following characteristics for staging: the extent of tumor penetration into the colon wall, with greater penetration generally correlating with a more dangerous tumor; the extent of invasion of the tumor through the colon wall and into other neighboring tissues, with greater invasion generally correlating with a more dangerous tumor; the extent of invasion of the tumor into the regional lymph nodes, with greater invasion generally correlating with a more dangerous tumor; and the extent of metastatic invasion into more distant tissues, such as the liver, with greater metastatic invasion generally correlating with a more dangerous disease state.

"Dukes A" and "Dukes B" colon cancers are neoplasias that have invaded into the wall of the colon but have not spread into other tissues. Dukes A colon cancers are cancers that have not invaded beyond the submucosa. Dukes B colon cancers are subdivided into two groups: "Dukes B1" and "Dukes B2". "Dukes B1" colon cancers are neoplasias that have invaded up to but not through the muscularis propria. Dukes B2 colon cancers are cancers that have breached completely through the muscularis propria. Over a five year period, patients with Dukes A cancer who receive surgical treatment (i.e. removal of the affected tissue) have a greater than 90% survival rate. Over the same period, patients with Dukes B1 and Dukes B2 cancer receiving surgical treatment have a survival rate of about 85% and 75%, respectively. Dukes A, B1 and B2 cancers are also referred to as T1, T2 and T3-T4 cancers, respectively.

"Dukes C" colon cancers are cancers that have spread to the regional lymph nodes, such as the lymph nodes of the gut. Patients with Dukes C cancer who receive surgical treatment alone have a 35% survival rate over a five year period, but this survival rate is increased to 60% in patients that receive chemotherapy.

"Dukes D" colon cancers are cancers that have metastasized to other organs. The liver is the most common organ in which metastatic colon cancer is found. Patients with Dukes D colon cancer have a survival rate of less than 5% over a five year period, regardless of the treatment regimen.

As noted above, early detection of colon neoplasia, coupled with appropriate intervention, is important for increasing patient survival rates. Present systems for screening for colon neoplasia are deficient for a variety of reasons, including a lack of specificity or sensitivity (e.g. Fecal Occult Blood Test, flexible sigmoidoscopy) or a high

cost and intensive use of medical resources (e.g. colonoscopy). Alternative systems for detection of colon neoplasia would be useful in a wide range of other clinical circumstances as well. For example, patients who receive surgical or pharmaceutical therapy for colon cancer may experience a relapse. It would be advantageous to have an alternative system for determining whether such patients have a recurrent or relapsed colon neoplasia. As a further example, an alternative diagnostic system would facilitate monitoring an increase, decrease or persistence of colon neoplasia in a patient known to have a colon neoplasia. A patient undergoing chemotherapy may be monitored to assess the effectiveness of the therapy.

Accordingly, in certain embodiments, the invention provides molecular markers that distinguish between cells that are not part of a colon neoplasia, referred to herein as "healthy cells", and cells that are part of a colon neoplasia (e.g. an adenoma or a colon cancer), referred to herein as "colon neoplasia cells". Certain molecular markers of the invention, including ColoUp1 and ColoUp2, are expressed at significantly higher levels in adenomas, Dukes A, Dukes B1, Dukes B2 and metastatic colon cancer of the liver (liver metastases) than in healthy colon tissue, healthy liver or healthy colon muscle. Certain molecular markers, including ColoUp1 and ColoUp2 are expressed at significantly higher levels in cell lines derived from colon cancer or cell lines engineered to imitate an aspect of a colon cancer cell. Particularly preferred molecular markers of the invention are markers that distinguish between healthy cells and cells of an adenoma. While not wishing to be bound to theory, it is contemplated that because adenomas are thought to be an obligate precursor for greater than 90% of colon cancers, markers that distinguish between healthy cells and cells of an adenoma are particularly valuable for screening apparently healthy patients to determine whether the patient is at increased risk for (predisposed to) developing a colon cancer. Furthermore, particularly preferred molecular markers are those that are actually present in the serum of an animal having a colon neoplasia, and in general, a secreted protein will generally occur in the serum only if it is secreted from a cell contacting a blood vessel, or a compartment in diffusional contact with a blood vessel. For example, protein secreted from a large or advanced colon cancer will generally be found in the blood stream, but a protein secreted from a colon adenoma may not be present in the blood unless it is secreted from the basolateral

face of the cell. Molecular markers that occur in the urine are generally derived from a polypeptide that is present in the blood. Optionally, a molecular marker is one that is present in the lumen of the colon (e.g., may be found in the intestinal mucous or in stool samples), and such a marker will generally be one that is secreted from the apical face of a cell.

In certain embodiments, the invention provides methods for using ColoUp molecular markers for determining whether a patient has or does not have a condition characterized by increased expression of one or more ColoUp nucleic acids or proteins described herein. In certain embodiments, the invention provides methods for determining whether a patient is or is not likely to have a colon neoplasia. In further embodiments, the invention provides methods for determining whether the patient is having a relapse or determining whether a patient's colon neoplasia is responding to treatment.

3. Methods for Identifying Candidate Molecular Markers for Colon Neoplasia

In certain aspects, the invention relates to the observation that when gene expression data is analyzed using carefully selected criteria, the likelihood of identifying strong candidate molecular markers of a colon neoplasia is quite high. Accordingly, in certain embodiments, the invention provides methods and criteria for analyzing gene expression data to identify candidate molecular markers for colon neoplasia. Although methods and criteria of the invention may be applied to essentially any relevant gene expression data, the benefits of using the inventive methods and criteria are readily apparent when applied to the copious data produced by highly parallel gene expression measurement systems, such as microarray systems. The human genome is estimated to be capable of producing roughly 20,000 to 100,000 different gene transcripts, thousands of which may show a change in expression level in healthy cells versus colon neoplasia cells. It is relatively cost-effective to obtain large quantities of gene expression data and to use this data to identify thousands of candidate molecular markers. However, a significant amount of labor intensive experimentation is generally needed to move from the identification of a candidate molecular marker to an effective diagnostic test for a health condition of interest. In fact, as of the time of filing of this application, the

resources required to generate a diagnostic test from a single candidate molecular marker identified by gene expression data are large enough that it is essentially impossible to extract commercially valuable and clinically useful diagnostics from a list of hundreds or thousands of genes whose expression levels change in a particular situation. Accordingly, there is a substantial practical value in being able to select a small number (e.g. ten or fewer) of high-quality molecular markers for further study.

In certain embodiments, candidate molecular markers for colon neoplasia may be selected by comparing gene expression in liver metastatic colon cancer samples ("liver mets"), normal (non-neoplastic) colon samples and normal liver samples. In this embodiment, candidate molecular markers are those genes (and their gene products) that have a level of expression in liver mets (assessed as a median expression level across the sample set) that is at least four times greater than the level of expression in normal colon samples (also assessed as a median expression level across the sample set). Furthermore, in this embodiment, the median level of expression in liver mets should be greater than the median level of expression in normal liver samples. The criteria employed in this embodiment provide a high threshold to eliminate most lower quality markers and further eliminate contaminants from liver tissue.

In certain embodiments, candidate molecular markers for colon neoplasia may be selected by comparing gene expression in normal colon to gene expression in a plurality of different cell lines cultured from metastatic colon cancer samples. For example median metastatic colon cancer cell line gene expression may be calculated as the median of 8 colon cancer cell lines of the Vaco colon cancer cell line series (Markowitz, S. et al. Science. 268: 1336-1338, 1995), such as the following liver metastatses-derived cell lines: V394, V576, V241, V9M, V400, V10M, V503, V786. In embodiments employing this criterion, candidate molecular markers are those genes (and their gene products) that have at least a three-fold higher median level of expression across the cell lines tested than in the normal colon tissue.

In certain embodiments, candidate molecular markers for colon neoplasia may be selected by comparing gene expression in normal colon to gene expression in a plurality of colon cancer xenografts grown in athymic mice ("xenografts"). In embodiments employing this criterion, candidate molecular markers are those genes (and their gene

products) that have at least a four-fold higher median level of expression across the xenografts tested than in the normal colon tissue.

In certain embodiments, candidate molecular markers for colon neoplasia may be selected by comparing maximum gene expression in normal colon to minimum gene expression in liver mets. In these embodiments, candidate molecular markers are those genes (and their gene products) that have a minimum gene expression in liver mets that is at least equal to the maximum gene expression in normal colon. Furthermore, in this embodiment, the median level of expression in liver mets should be greater than the median level of expression in normal liver samples.

In a preferred embodiment, a list of candidate molecular markers for colon neoplasia is selected by first identifying a subset of genes having a four-fold greater median expression in liver mets that in normal colon and in normal liver. This subset is then further narrowed to a final list by identifying those genes that have a three-fold greater median expression across colon cancer cell lines than in normal colon. Optionally, a particularly preferred list may be generated by further selecting those genes having a minimum gene expression in liver mets that is greater than or equal to the maximum gene expression in normal colon. The gene products (e.g. proteins and nucleic acids) of the short list of genes generated in these preferred embodiments constitute a list of high-quality candidate molecular markers for colon cancer.

In another preferred embodiment, a list of candidate molecular markers for colon neoplasia is selected by first identifying a subset of genes having a four-fold greater median expression in liver mets that in normal colon and in normal liver. This subset is then further narrowed by identifying those genes that have a nine-fold greater median expression in liver mets than in normal colon. This subset is then further narrowed to a final list by identifying those genes that have a four-fold greater median expression across colon cancer cell lines than in normal colon. The gene products (e.g. proteins and nucleic acids) of the short list of genes generated in these preferred embodiments constitute a list of high-quality candidate molecular markers for colon cancer.

Depending on the nature of the intended use for the molecular marker it may be desirable to add further criteria to any of the preceding embodiments. In certain embodiments, the invention relates to candidate molecular markers for categorizing a

patient as likely to have or not likely to have a colon neoplasia (including adenomas and colon cancers), and in these embodiments, a high-quality candidate molecular marker will be expressed from a gene having an increased expression in both adenomas and liver mets relative to normal colon, and preferably in other colon cancer stages, including Dukes A, Dukes B1, Dukes B2 and Dukes C. In certain embodiments the invention relates to candidate molecular markers for categorizing a patient as likely to have or not likely to have a colon cancer (including metastatic and non-metastatic forms), and in these embodiments, a high-quality candidate molecular marker will be expressed from a gene having an increased expression in liver mets relative to adenomas and normal colon, and preferably there will be elevated expression in other colon cancer stages, including Dukes A, Dukes B1, Dukes B2 and Dukes C. In certain embodiments, the invention relates to candidate molecular markers for categorizing a patient as likely or not likely to have a metastatic colon cancer, and in such embodiments, a comparison to gene expression in other colon neoplasias (e.g. adenomas, Dukes A, Dukes B1, Dukes B2, Dukes C), while potentially useful, is not necessary, although it is noted that expression in non-metastatic states may indicate that a candidate molecular marker is not of high quality for distinguishing metastatic colon cancer from non-metastatic states.

Furthermore, in those embodiments pertaining to molecular markers to be used for detection in a body fluid, such as blood, a high quality molecular marker will preferably be a secreted protein. In those embodiments pertaining to neoplasia identification or targeting, a high quality molecular marker will preferably be a protein with a portion adherent to and exposed on the extracellular surface of a neoplasia, such as a transmembrane protein with a significant extracellular portion.

Gene expression data may be gathered using one or more of the many known and appropriate techniques that, in view of this specification, may be selected to one of skill in the art. In certain preferred embodiments, gene expression data is gathered by a highly parallel system, meaning a system that allows simultaneous or near-simultaneous collection of expression data for one hundred or more gene transcripts. Exemplary highly parallel systems include probe arrays ("arrays") that are often divided into microarrays and macroarrays, where microarrays have a much higher density of individual probe species per area. Arrays generally consist of a surface to which probes that correspond in

sequence to gene products (e.g., cDNAs, mRNAs, oligonucleotides) are bound at known positions. The probes can be, e.g., a synthetic oligomer, a full-length cDNA, a less-than full length cDNA, or a gene fragment. Usually a microarray will have probes corresponding to at least 100 gene products and more preferably, 500, 1000, 4000 or more. Probes may be small oligomers or larger polymers, and there may be a plurality of overlapping or non-overlapping probes for each transcript.

The nucleic acids to be contacted with the microarray may be prepared in a variety of ways. Methods for preparing total and poly(A)+ RNA are well known and are described generally in Sambrook et al., supra. Labeled cDNA may be prepared from mRNA by oligo dT-primed or random-primed reverse transcription, both of which are well known in the art (see e.g., Klug and Berger, 1987, Methods Enzymol. 152:316-325). cDNAs may be labeled by incorporation of labeled nucleotides or by labeling after synthesis. Preferred labels are fluorescent labels.

Nucleic acid hybridization and wash conditions are chosen so that the population of labeled nucleic acids will specifically hybridize to appropriate, complementary probes affixed to the matrix. Optimal hybridization conditions will depend on the length (e.g., oligomer versus polynucleotide greater than 200 bases) and type (e.g., RNA, DNA, PNA) of labeled nucleic acids and immobilized polynucleotide or oligonucleotide. General parameters for specific (i.e., stringent) hybridization conditions for nucleic acids are described in Sambrook et al., supra, and in Ausubel et al., 1987, Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience, New York, which is incorporated in its entirety for all purposes. Non-specific binding of the labeled nucleic acids to the array can be decreased by treating the array with a large quantity of non-specific DNA -- a so-called "blocking" step.

Signals, such as fluorescent emissions for each location on an array are generally recorded, quantitated and analyzed using a variety of computer software. Signal for any one gene product may be normalized by a variety of different methods. Arrays preferably include control and reference probes. Control probes are nucleic acids which serve to indicate that the hybridization was effective. Reference probes allow the normalization of results from one experiment to another, and to compare multiple experiments on a quantitative level. Reference probes are typically chosen to correspond

to genes that are expressed at a relatively constant level across different cell types and/or across different culture conditions. Exemplary reference nucleic acids include housekeeping genes of known expression levels, e.g., GAPDH, hexokinase and actin.

Following the data gathering operation, the data will typically be reported to a data analysis system. To facilitate data analysis, the data obtained by the reader from the device will typically be analyzed using a digital computer. Typically, the computer will be appropriately programmed for receipt and storage of the data from the device, as well as for analysis and reporting of the data gathered, e.g., subtraction of the background, deconvolution multi-color images, flagging or removing artifacts, verifying that controls have performed properly, normalizing the signals, interpreting fluorescence data to determine the amount of hybridized target, normalization of background and single base mismatch hybridizations, and the like. Various analysis methods that may be employed in such a data analysis system, or by a separate computer are described herein.

A number of methods for constructing or using arrays are described in the following references. Schena et al., 1995, Science 270:467-470; DeRisi et al., 1996, Nature Genetics 14:457-460; Shalon et al., 1996, Genome Res. 6:639-645; Schena et al., 1995, Proc. Natl. Acad. Sci. USA 93:10539-11286; Fodor et al., 1991, Science 251:767-773; Pease et al., 1994, Proc. Natl. Acad. Sci. USA 91:5022-5026; Lockhart et al., 1996, Nature Biotech 14:1675; U.S. Pat. Nos. 6,051,380; 6,083,697; 5,578,832; 5,599,695; 5,593,839; 5,631,734; 5,556,752; 5,510,270; EP No. 0 799 897; PCT No. WO 97/29212; PCT No. WO 97/27317; EP No. 0 785 280; PCT No. WO 97/02357; EP No. 0 728 520; EP No. 0 721 016; PCT No. WO 95/22058.

A variety of companies provide microarrays and software for extracting certain information from microarray data. Such companies include Affymetrix (Santa Clara, CA), GeneLogic (Gaithersburg, MD) and Eos Biotechnology Inc. (South San Francisco, CA).

While the above discussion focuses on the use of arrays for the collection of gene expression data, such data may also be obtained through a variety of other methods, that, in view of this specification, are known to one of skill in the art. Such methods include the serial analysis of gene expression (SAGE) technique, first described in Velculescu et al. (1995) *Science* 270, 484-487. Reverse transcriptase – polymerase chain reaction (RT-

PCR) may be used, and particularly in combination with fluorescent probe systems such as the Taqman™ fluorescent probe system. Numerous RT-PCR samples can be analyzed simultaneously by conducting parallel PCR amplification, e.g., by multiplex PCR. Further techniques include dotblot analysis and related methods (*see, e.g.*, G. A. Beltz et al., in Methods in Enzymology, Vol. 100, Part B, R. Wu, L. Grossmam, K. Moldave, Eds., Academic Press, New York, Chapter 19, pp. 266-308, 1985), Northern blots and *in situ* hybridization (probing a tissue sample directly).

The quality and biological relevance of gene expression data will be significantly affected by the quality of the biological material used to obtain gene expression. In preferred embodiments, the methods described herein for identifying candidate molecular markers for colon neoplasia employ tissue samples obtained with appropriate consent from human patients and rapidly frozen. At a point prior to gene expression analysis, the tissue sample is preferably prepared by carefully dissecting away as much heterogeneous tissue as is possible with the available tools. In other words, for a colon cancer sample, adherent non-cancerous tissue should be dissected away, to the extent that it is possible. In preferred embodiments, healthy tissue is obtained from a subject that has a colon neoplasia but is tissue that is not directly entangled in a neoplasia.

Example 1, below, illustrates the operation of a method of selecting high-quality molecular markers, and the following markers were selected, using criteria disclosed herein, from microarray expression data: ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7 and ColoUp8. In addition, osteopontin was identified as having expression characteristics very similar to those identified using the selection criteria. Further experimentation (see Examples) demonstrated that these molecular markers fall into four categories: "secreted" (ColoUp1, ColoUp2 and osteopontin), "transmembrane" (ColoUp3), "transcription factors" (ColoUp4, ColoUp5) and "other" (ColoUp6, ColoUp7, ColoUp8). Further experimentation also demonstrated that ColoUp1, ColoUp2, ColoUp3, ColoUp5 and ColoUp7 are, generally speaking, expressed at higher levels in a variety of colon neoplasias (adenomas, Dukes B tumors, Dukes C tumors and liver mets) than in healthy cells. In addition, further experimentation demonstrated that osteopontin is overexpressed in colon cancers (Dukes B, Dukes C and liver mets) relative to adenomas and normal colon.

In certain embodiments, a preferred molecular marker for use in a diagnostic test that employs a body fluid sample, such as a blood or urine sample, or an excreted sample material, such as stool, is a secreted protein, such as the secreted portion of a ColoUp1 protein, ColoUp2 protein or osteopontin protein.

In certain embodiments, a preferred molecular marker for a method that involves targeting or marking a colon neoplasia is a transmembrane protein, such as ColoUp3, and particularly the extracellular portion of ColoUp3. Transmembrane proteins are desirable for such methods because they are both anchored to the neoplastic cell and exposed to the extracellular surface.

In certain embodiments, a preferred molecular marker for use in a diagnostic test to distinguish subjects likely to have a colon neoplasia from those not likely to have a colon neoplasia is a gene product of the ColoUp1, ColoUp2, ColoUp3, ColoUp4 or ColoUp5 genes. Examples of suitable gene products include proteins, both secreted and not secreted and transcripts. In embodiments employing proteins that are not secreted, such as ColoUp3, ColoUp4 and ColoUp5, a preferred embodiment of the diagnostic test is a test for the presence of the protein or transcript in cells shed from the colon or colon neoplasia (which, in the case of metastases is not necessarily located in the colon) into a sample material, such as stool. In embodiments employing proteins that are secreted, such as ColoUp1 and ColoUp2, a preferred embodiment of the diagnostic test is a test for the presence of the protein in a body fluid, such as urine or blood or an excreted material, such as stool. It should be noted, however, that intracellular protein may be present in a body fluid if there is significant cell lysis or through some other process. Likewise, secreted proteins are likely to be adherent, even if at a relatively low level, to the cells in which they were produced.

In certain embodiments, a preferred molecular marker for distinguishing subjects having a colon cancer from those having an adenoma or a normal colon is gene product of the ColoUp6 and osteopontin genes. In embodiments preferably employing marker proteins that are secreted, such as a test using a body fluid sample, a preferred marker is a secreted osteopontin protein.

<u>ColoUp1:</u>

A human ColoUp1 nucleic acid sequence encodes a full-length protein of 1361 amino acids. SignalP V1.1 predicts that human ColoUp1 protein has an N-terminal signal peptide that is cleaved between either amino acids 30-31(ATS-TV) or amino acids 33-34 (TVA-AG). Four potential glycosylation sites are identified in ColoUp1 protein. Further, ColoUp1 protein is predicted to have multiple serine, threonine, and tyrosine phosphorylation sites for kinases such as protein kinase C, cAMP- and cGMP-dependent protein kinases, casein kinase II, and tyrosine kinases. The ColoUp1 protein shares limited sequence homology to a human transmembrane protein 2 (See Scott et al. 2000 Gene 246:265-74). A mouse ColoUp1 homolog is identified in existing GenBank databases and is linked with mesoderm development (see Wines et al. 2001 Genomics. 88-98; GenBank entry AAG41062, AY007815 for the 1179 bp nucleic acid sequence entry, with 363/390 (93%) identities with human ColoUp1).

As demonstrated herein, ColoUp1 is secreted from both the basolateral and apical surfaces of intestinal cells.

ColoUp2:

The ColoUp2 nucleic acid sequence encodes a full-length protein of 755 amino acids. The application also discloses certain polymorphisms that have been observed, for example at nucleotide 113 GCC→ACC (Ala-Thr); nt 480 GAA→GGA (Glu-Gly); and at nt 2220 CAG→CGG (Gln-Arg). The sequence of ColoUp2 protein is similar to that of alpha 3 type VI collagen, isoform 2 precursor. In addition, a few domains are identified in the ColoUp2 protein such as a von Willebrand factor type A domain (vWF) and an EGF-like domain. The vWF domain is found in various plasma proteins such as some complement factors, the integrins, certain collagen, and other extracellular proteins. Proteins with vWF domains participate in numerous biological events which involve interaction with a large array of ligands, for example, cell adhesion, migration, homing, pattern formation, and signal transduction. The EGF-like domain consisting of about 30-40 amino acid residues has been found many proteins. The functional significance of EGF domains is not yet clear. However, a common feature is that these EGF-like repeats are found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted.

As demonstrated herein, ColoUp2 is secreted from both the apical and basolateral surfaces of intestinal cells, and can be found in the blood in two different forms, a full-length secreted form and a C-terminal fragment (approximately 55 kDa).

Osteopontin:

The Osteopontin nucleic acid sequence encodes a full-length protein of 300 amino acids. Osteopontin is an acidic glycoprotein and is produced primarily by osteoclasts, macrophages, T-cells, kidneys, and vascular smooth muscle cells. As a cytokine, Osteopontin is known to contribute substantially to metastasis formation by various cancers. In addition, it contributes to macrophage homing and cellular immunity, mediates neovascularization, inhibits apoptosis, and maintains the homeostasis of free calcium (see a review, Weber GF. 2001 Biochim Biophys Acta. 1552:61-85).

ColoUp3:

The ColoUp3 nucleic acid sequence encodes a full-length protein of 829 amino acids. ColoUp3 is referred to in the literature as P-cadherin (or cadherin 3, type 1). P-cadherin belongs to a cadherin family that includes E- cadherin and N-cadherin. P-cadherin is expressed in placenta and stratified squamous epithelia (see Shimoyama et al. 1989 J Cell Biol. 109:1787–94), but not in normal colon. P-cadherin null mice develop mammary gland hyperplasia, dysplasia, and abnormal lymphoid infiltration (see Radice et al. 1997 J Cell Biol. 139:1025–32), demonstrating that loss of normal P-cadherin expression leads to cellular and glandular abnormalities. It has been shown that P-cadherin is aberrantly expressed in inflamed and dysplastic colitic mucosa, with concomitant E-cadherin downregulation. Recently, aberrant P-cadherin expression is found as an early event in hyperplastic and dysplastic transformation in the colon (see Hardy et al. 2002 Gut. 50:513-514).

ColoUp4:

The ColoUp4 nucleic acid sequence encodes a full-length protein of 694 amino acids. ColoUp4 is referred to in the literature as NF-E2 related factor 3 (NRF3). NRF3 was identified and characterized as a novel Cap'n' collar (CNC) factor, with a basic

region-leucine zipper domain highly homologous to those of other CNC proteins such as NRF1 and NRF2. These CNC factors bind to Maf recognition elements (MARE) through heterodimer formation with small Maf proteins *In vitro* and *in vivo* analyses showed that NRF3 can heterodimerize with MafK and that this complex binds to the MARE in the chicken β-globin enhancer and can activate transcription. NRF3 mRNA is highly expressed in human placenta and B cell and monocyte lineage. (see Kobayashi et al. 1999 J Biol Chem. 274:6443-52).

ColoUp5:

The ColoUp5 nucleic acid sequence encodes a full-length protein of 402 amino acids. ColoUp5 is referred to in the literature as FoxQ1 (Forkhead box, subclass q, member 1, formerly known as HFH-1). FoxQ1 is a member of the evolutionarily conserved winged helix/forkhead transcription factor gene family. The hallmark of this family is a conserved DNA binding region of approximately 110 amino acids (FOX domain). Members of the FOX gene family are found in a broad range of organisms from yeast to human. Human FoxQ1 gene is expressed in different tissues such as stomach, trachea, bladder, and salivary gland. FoxQ1 gene plays important roles in tissue-specific gene regulation and development, for example, embryonic development, cell cycle regulation, cell signaling, and tumorigenesis. The FoxQ1 gene is located on chromosome 6p23-25. Sequence analysis indicates that human FoxQ1 shows 82% homology with the mouse Foxq1 gene (formerly Hfh-1L) and with a revised sequence of the rat FoxQ1 gene (formerly Hfh-1). Mouse FoxQ1 was shown to regulate differentiation of hair in Satin mice. The DNA-binding motif (i.e., the FOX domain) is well conserved, showing 100% identity in human, mouse, and rat. The human FoxQ1 protein sequence contains two putative transcriptional activation domains, which share a high amino acid identity with the corresponding mouse and rat domains (see Bieller et al. 2001 DNA Cell Biol. 20:555-61).

ColoUp6:

The ColoUp6 nucleic acid sequence encodes a full-length protein of 209 amino acids. The ColoUp6 protein is 99% identical to the C-terminal portion of keratin 23 (or

cytokeratin 23, or the type I intermediate filament cytokeratin), and accordingly the term ColoUp6 includes both the 209 amino acid protein (and related nucleic acids, fragments, variants, etc.) and the cytokeratin 23 amino acid sequence of GenBank entry BAA92054.1 (and related nucleic acids, fragments, variants, etc.). Keratin 23 mRNA was found highly induced in different pancreatic cancer cell lines in response to sodium butyrate. The keratin 23 protein has 422 amino acids, and has an intermediate filament signature sequence and extensive homology to type I keratins. It is suggested that keratin 23 is a novel member of the acidic keratin family that is induced in pancreatic cancer cells undergoing differentiation by a mechanism involving histone hyperacetylation (See Zhang et al. 2001 Genes Chromosomes Cancer. 30:123-35).

ColoUp7:

The ColoUp7 nucleic acid sequence is an EST sequence. No information relating to the function of the ColoUp7 gene is identified.

ColoUp8:

The ColoUp8 nucleic acid sequence encodes a full-length protein of 278 amino acids. No function has been suggested relating to the ColoUp8 gene.

Accordingly, in certain embodiments, the application provides isolated, purified or recombinant ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7, ColoUp8 and osteopontin nucleic acids. In certain embodiments, such nucleic acids may encode a complete or partial ColoUp polypeptide or such nucleic acids may also be probes or primers useful for methods involving detection or amplification of ColoUp nucleic acids. In certain embodiments, a ColoUp nucleic acid is single-stranded or double-stranded and composed of natural nucleic acids, nucleotide analogs, or mixtures thereof. In certain embodiments, the application provides isolated, purified or recombinant nucleic acids comprising a nucleic acid sequence that is at least 90% identical to a nucleic acid sequence of any of SEQ ID Nos: 3-12, or a complement thereof, and optionally at least 95%, 97%, 98%, 99%, 99.3%, 99.5%, 99.7% or 100% identical to a nucleic acid of any of SEQ ID Nos: 3-12, or a complement thereof. In

certain preferred embodiments, the application provides a isolated, purified or recombinant nucleic acids comprising a nucleic acid sequence that is at least 90%, 95%, 97%, 98%, 99%, 99.3%, 99.5%, 99.7% or 100% identical to a nucleic acid of any of SEQ ID Nos: 3-12, or a complement thereof. In certain embodiments, the application provides isolated, purified or recombinant nucleic acids comprising a nucleic acid sequence that encodes a polypeptide that is at least 90% identical to an amino acid sequence of any of SEQ ID Nos: 1-3 or 13-21, or a complement thereof, and optionally at least 95%, 97%, 98%, 99%, 99.3%, 99.5%, 99.7% or 100% identical to an amino acid sequence of any of SEQ ID Nos: 1-3 or 13-21, or a complement thereof. In certain preferred embodiments, the application provides isolated, purified or recombinant nucleic acids comprising a nucleic acid sequence that encodes a polypeptide that is at least 90% identical to an amino acid sequence of any of SEQ ID Nos: 3, 14 or 21, or a complement thereof, and optionally at least 95%, 97%, 98%, 99%, 99.3%, 99.5%, 99.7% or 100% identical to an amino acid sequence of any of SEQ ID Nos: 3, 14 or 21, or a complement thereof.

In further embodiments, the application provides expression constructs, vectors and cells comprising a ColoUp nucleic acid. Expression constructs are nucleic acid constructs that are designed to permit expression of an expressible nucleic acid (e.g. a ColoUp nucleic acid) in a suitable cell type or in vitro expression system. A variety of expression construct systems are, in view of this specification, well known in the art, and such systems generally include a promoter that is operably linked to the expressible nucleic acid. The promoter may be a constitutive promoter, as in the case of many viral promoters, or the promoter may be a conditional promoter, as in the case of the prokaryotic lacI-repressible, IPTG-inducible promoter and as in the case of the eukaryotic tetracycline-inducible promoter. Vectors refer to any nucleic acid that is capable of transporting another nucleic acid to which it has been linked between different cells or viruses. One type of vector is an episome, i.e., a nucleic acid capable of extra-chromosomal replication, such as a plasmid. Episome-type vectors typically carry an origin of replication that directs replication of the vector in a host cell. Another type of vector is an integrative vector that is designed to recombine with the genetic material of a host cell. Vectors may be both autonomously replicating and integrative, and the properties of a vector may differ depending on the cellular context (i.e. a vector may be

autonomously replicating in one host cell type and purely integrative in another host cell type). Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". Vectors that carry an expression construct are generally expression vectors. Vectors have been designed for a variety of cell types. For example, in the bacterium E. coli, commonly used vectors include pUC plasmids, pBR322 plasmids, pBlueScript and M13 plasmids. In insect cells (e.g. SF-9, SF-21 and High-Five cells), commonly used vectors include BacPak6 (Clontech) and BaculoGold (Pharmingen) (both Clontech and Pharmingen are divisions of Becton, Dickinson and Co., Franklin Lakes, New Jersey). In mammalian cells (e.g. Chinese hamster ovary (CHO) cells, Vaco cells and human embryonic kidney (HEK) cells), commonly used vectors include pCMV vectors (Stratagene, Inc., La Jolla, California), and pRK vectors. In certain embodiments, the application provides cells that comprise a ColoUp nucleic acid, particularly a recombinant ColoUp nucleic acid, such as an expression construct or vector that comprises a ColoUp nucleic acid. Cells may be eukaryotic or prolaryotic, depending on the anticipated use. Prokaryotic cells, especially E. coli, are particularly useful for storing and replicating nucleic acids, particularly nucleic acids carried on plasmid or viral vectors. Bacterial cells are also particularly useful for expressing nucleic acids to produce large quantities of recombinant protein, but bacterial cells do not usually mimic eukaryotic post-translational modifications, such as glycosylations or lipid-modifications, and so will tend to be less suitable for production of proteins in which the post-translational modification state is significant. Eukaryotic cells, and especially cell types such as insect cells that work with baculovirus-based protein expression systems, and Chinese hamster ovary cells, are good systems for expressing eukaryotic proteins that have significant post-translational modifications. Eukaryotic cells are also useful for studying various aspects of the function of eukaryotic proteins. For example, colon cancer cell lines are good model systems for studying the role of ColoUp genes and proteins in colon cancers.

In certain aspects the application further provides methods for preparing ColoUp polypeptides. In general, such methods comprise obtaining a cell that comprises a nucleic acid encoding a ColoUp polypeptide, and culturing the cell under conditions that cause production of the ColoUp polypeptide. Polypeptides produced in this manner may

be obtained from the appropriate cell or culture fraction. For example, secreted proteins are most readily obtained from the culture supernatant, soluble intracellular proteins are most readily obtained from the soluble fraction of a cell lysate, and membrane proteins are most readily obtained from a membrane fraction. However, proteins of each type can generally be found in all three types of cell or culture fraction. Crude cellular or culture fractions may be subjected to further purification procedures to obtain substantially purified ColoUp polypeptides. Common purification procedures include affinity purification (e.g. with hexahistidine-tagged polypeptides), ion exchange chromatography, reverse phase chromatography, gel filtration chromatography, etc.

In certain aspects the application provides recombinant, isolated, substantially purified or purified ColoUp1, ColoUp2, ColoUp3, ColoUp4, ColoUp5, ColoUp6, ColoUp7, ColoUp8 and osteopontin polypeptides. In certain embodiments, such polypeptides may encode a complete or partial ColoUp polypeptide. In certain embodiments, a ColoUp polypeptide is composed of natural amino acids, amino acid analogs, or mixtures thereof. ColoUp polypeptides may also include one or more post-translational modifications, such as glycosylation, phosphorylation, lipid modification, acetylation, etc. In certain embodiments, the application provides isolated, substantially purified, purified or recombinant polypeptides comprising an amino acid sequence that is at least 90% identical to an amino acid sequence of any of SEQ ID Nos: 1-3 or 13-21 and optionally at least 95%, 97%, 98%, 99%, 99.3%, 99.5% or 99.7% identical to a nucleic acid of any of SEQ ID Nos: 1-3 or 13-21. In certain preferred embodiments, the application provides a isolated, substantially purified, purified or recombinant polypeptide comprising an amino acid sequence that is at least 90%, 95%, 97%, 98%, 99%, 99.3%, 99.5% or 99.7% identical to a nucleic acid of any of SEQ ID Nos: 3, 14 or 21. In certain preferred embodiments, the application provides an isolated, subtstantially purified, purified or recombinant polypeptide comprising an amino acid sequence that differs from SEQ ID Nos. 3, 14 or 21 by no more than 4 amino acid substitutions, additions or deletions. Optionally, a polypeptide of the invention comprises an additional moiety, such as an additional polypeptide sequence or other added compound, with a particular function, such as an epitope tag that facilitates detection of the recombinant polypeptide with an antibody, a purification moiety that facilitates purification (e.g. by

affinity purification), a detection moiety, that facilitates detection of the polypeptide in vivo or in vitro, or an antigenic moiety that increases the antigenicity of the polypeptide so as to facilitate antibody production. Often, a single moiety will provide multiple functionalities. For example, an epitope tag will generally also assist in purification, because an antibody that recognizes the epitope can be used in an affinity purification procedure as well. Examples of commonly used epitope tags are: an HA tag, a hexahistidine tag, a V5 tag, a Glu-Glu tag, a c-myc tag, a VSV-G tag, a FLAG tag, an enterokinase cleavage site tag and a T7 tag. Commonly used purification moieties include: a hexahistidine tag, a glutathione-S-transferase domain, a cellulose binding domain and a biotin tag. Commonly used detection moieties include fluorescent proteins (e.g. green fluorescent proteins), a biotin tag, and chromogenic/fluorogenic enzymes (e.g. beta-galactosidase and luciferase). Commonly used antigenic moieties include the keyhole limpet hemocyanin and serum albumins. Note that these moieties need not be polypeptides and need not be connected to the polypeptide by a traditional peptide bond.

4. Antibodies and Uses Therefor

Another aspect of the invention pertains to an antibody specifically reactive with a ColoUp polypeptide that is effective for decreasing a biological activity of the polypeptide, preferably antibodies that are specifically reactive with ColoUp polypeptides such as ColoUp1 and ColoUp2 polypeptides. For example, by using immunogens derived from a ColoUp polypeptide, e.g., based on the cDNA sequences, anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, Antibodies: A Laboratory Manual ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal, such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide (e.g., a ColoUp polypeptide or an antigenic fragment which is capable of eliciting an antibody response, or a fusion protein). Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of a ColoUp polypeptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as

antigen to assess the levels of antibodies. In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of a ColoUp polypeptide of a mammal, e.g., antigenic determinants of a protein set forth in SEQ ID Nos: 1-3 and 13-21, more preferably SEQ ID Nos: 1-3 or 21.

In one embodiment, antibodies are specific for the secreted proteins as encoded by nucleic acid sequences as set forth in SEQ ID Nos: 4-5. In another embodiment, the antibodies are immunoreactive with one or more proteins having an amino acid sequence that is at least 80% identical to an amino acid sequence as set forth in SEQ ID Nos: 1-3 and 13-21, preferably SEQ ID Nos: 1-3 or 21. In other embodiments, an antibody is immunoreactive with one or more proteins having an amino acid sequence that is at least 85%, 90%, 95%, 98%, 99%, 99.3%, 99.5%, 99.7% identical or 100% identical to an amino acid sequence as set forth in SEQ ID Nos: 1-3 and 13-21. More preferably, the antibody is immunoreactive with one or more proteins having an amino acid sequence that is at least 85%, 90%, 95%, 98%, 99%, 99.3%, 99.5%, 99.7% or identical to an amino acid sequence as set forth in SEQ ID NOs: 1-3 or 21. In certain preferred embodiments, the invention provides an antibody that binds to an epitope including the C-terminal portion of the polypeptide of SEQ ID Nos: 3, 14 or 21. In certain preferred embodiments, the invention provides an antibody that binds to an epitope of a ColoUp2 polypeptide that is prevalent in the blood of an animal having a colon neoplasia, such SEQ ID No: 3 or 21.

Following immunization of an animal with an antigenic preparation of a ColoUp polypeptide, anti-ColoUp antisera can be obtained and, if desired, polyclonal anti-ColoUp antibodies can be isolated from the serum. To produce monoclonal antibodies, antibody-producing cells (lymphocytes) can be harvested from an immunized animal and fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Such techniques are well known in the art, and include, for example, the hybridoma technique (originally developed by Kohler and Milstein, (1975) Nature, 256: 495-497), the human B cell hybridoma technique (Kozbar et al., (1983) Immunology Today, 4: 72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., (1985) Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. pp. 77-96). Hybridoma cells can be screened

immunochemically for production of antibodies specifically reactive with a mammalian ColoUp polypeptide of the present invention and monoclonal antibodies isolated from a culture comprising such hybridoma cells. In one embodiment anti-human ColoUp antibodies specifically react with the protein encoded by a nucleic acid having SEQ ID Nos: 4-12; more preferably the antibodies specifically react with the protein encoded by a nucleic acid having SEQ ID Nos: 4 or 5, and preferably a secreted protein that is produced by the expression of a nucleic acid having a sequence of SEQ ID Nos: 4 or 5.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with one of the subject ColoUp polypeptides. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab)_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab)_2$ fragment can be treated to reduce disulfide bridges to produce Fab fragments. The antibody of the present invention is further intended to include bispecific, single-chain, and chimeric and humanized molecules having affinity for a ColoUp polypeptide conferred by at least one CDR region of the antibody. In preferred embodiments, the antibodies, the antibody further comprises a label attached thereto and able to be detected, (e.g., the label can be a radioisotope, fluorescent compound, enzyme or enzyme co-factor).

In certain preferred embodiments, an antibody of the invention is a monoclonal antibody, and in certain embodiments the invention makes available methods for generating novel antibodies. For example, a method for generating a monoclonal antibody that binds specifically to a ColoUp polypeptide, such as a ColoUp2 polypeptide may comprise administering to a mouse an amount of an immunogenic composition comprising the ColoUp2 polypeptide effective to stimulate a detectable immune response, obtaining antibody-producing cells (e.g. cells from the spleen) from the mouse and fusing the antibody–producing cells with myeloma cells to obtain antibody-producing hybridomas, and testing the antibody-producing hybridomas to identify a hybridoma that produces a monocolonal antibody that binds specifically to the ColoUp2 polypeptide. Once obtained, a hybridoma can be propagated in a cell culture, optionally in culture conditions where the hybridoma-derived cells produce the monoclonal

antibody that binds specifically to the ColoUp2 polypeptide. The monoclonal antibody may be purified from the cell culture.

Anti-ColoUp antibodies can be used, e.g., to detect ColoUp polypeptides in biological samples and/or to monitor ColoUp polypeptide levels in an individual, for determining whether or not said patient is likely to develop colon cancer or is more likely to harbor colon adenomas, or allowing determination of the efficacy of a given treatment regimen for an individual afflicted with colon neoplasia, colon cancer, metastatic colon cancer and colon adenomas. The level of ColoUp polypeptide may be measured in a variety of sample types such as, for example, in cells , stools, and/or in bodily fluid, such as in whole blood samples, blood serum, blood plasma and urine. The adjective "specifically reactive with" as used in reference to an antibody is intended to mean, as is generally understood in the art, that the antibody is sufficiently selective between the antigen of interest (e.g. a ColoUp polypeptide) and other antigens that are not of interest that the antibody is useful for, at minimum, detecting the presence of the antigen of interest in a particular type of biological sample. In certain methods employing the antibody, a higher degree of specificity in binding may be desirable. For example, an antibody for use in detecting a low abundance protein of interest in the presence of one or more very high abundance protein that are not of interest may perform better if it has a higher degree of selectivity between the antigen of interest and other cross-reactants. Monoclonal antibodies generally have a greater tendency (as compared to polyclonal antibodies) to discriminate effectively between the desired antigens and cross-reacting polypeptides. In addition, an antibody that is effective at selectively identifying an antigen of interest in one type of biological sample (e.g. a stool sample) may not be as effective for selectively identifying the same antigen in a different type of biological sample (e.g. a blood sample). Likewise, an antibody that is effective at identifying an antigen of interest in a purified protein preparation that is devoid of other biological contaminants may not be as effective at identifying an antigen of interest in a crude biological sample, such as a blood or urine sample. Accordingly, in preferred embodiments, the application provides antibodies that have demonstrated specificity for an antigen of interest (particularly, although not limited to, a ColoUp1 or ColoUp2 polypeptide) in a sample type that is likely to be the sample type of choice for use of the

antibody. In a particularly preferred embodiment, the application provides antibodies that bind specifically to a ColoUp1 or ColoUp2 polypeptide in a protein preparation from blood (optionally serum or plasma) from a patient that has a colon neoplasia or that bind specifically in a crude blood sample (optionally a crude serum or plasma sample).

One characteristic that influences the specificity of an antibody:antigen interaction is the affinity of the antibody for the antigen. Although the desired specificity may be reached with a range of different affinities, generally preferred antibodies will have an affinity (a dissociation constant) of about 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9} or less.

In addition, the techniques used to screen antibodies in order to identify a desirable antibody may influence the properties of the antibody obtained. For example, an antibody to be used for certain therapeutic purposes will preferably be able to target a particular cell type. Accordingly, to obtain antibodies of this type, it may be desirable to screen for antibodies that bind to cells that express the antigen of interest (e.g. by fluorescence activated cell sorting). Likewise, if an antibody is to be used for binding an antigen in solution, it may be desirable to test solution binding. A variety of different techniques are available for testing antibody:antigen interactions to identify particularly desirable antibodies. Such techniques include ELISAs, surface plasmon resonance binding assays (e.g. the Biacore binding assay, Bia-core AB, Uppsala, Sweden), sandwich assays (e.g. the paramagnetic bead system of IGEN International, Inc., Gaithersburg, Maryland), western blots, immunoprecipitation assays and immunohistochemistry.

Another application of anti-ColoUp antibodies of the present invention is in the immunological screening of cDNA libraries constructed in expression vectors such as gt11, gt18-23, ZAP, and ORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, gt11 will produce fusion proteins whose amino termini consist of ß-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a ColoUp polypeptide, e.g., other orthologs of a particular protein or other paralogs from the same species, can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with the appropriate anti-ColoUp antibodies. Positive phage detected by this assay can then be

isolated from the infected plate. Thus, the presence of ColoUp homologs can be detected and cloned from other animals, as can alternate isoforms (including splice variants) from humans.

5. Methods for Detecting Molecular Markers in a Patient

In certain embodiments, the invention provides methods for detecting molecular markers, such as proteins or nucleic acid transcripts of the ColoUp markers described herein. In certain embodiments, a method of the invention comprises providing a biological sample and probing the biological sample for the presence of a ColoUp marker. Information regarding the presence or absence of the ColoUp marker, and optionally the quantitative level of the ColoUp marker, may then be used to draw inferences about the nature of the biological sample and, if the biological sample was obtained from a subject, the health state of the subject.

Samples for use with the methods described herein may be essentially any biological material of interest. For example, a sample may be a tissue sample from a subject, a fluid sample from a subject, a solid or semi-solid sample from a subject, a primary cell culture or tissue culture of materials derived from a subject, cells from a cell line, or medium or other extracellular material from a cell or tissue culture, or a xenograft (meaning a sample of a colon cancer from a first subject, e.g. a human, that has been cultured in a second subject, e.g. an immunocompromised mouse). The term "sample" as used herein is intended to encompass both a biological material obtained directly from a subject (which may be described as the primary sample) as well as any manipulated forms or portions of a primary sample. For example, in certain embodiments, a preferred fluid sample is a blood sample. In this case, the term sample is intended to encompass not only the blood as obtained directly from the patient but also fractions of the blood, such as plasma, serum, cell fractions (e.g. platelets, erythrocytes, lymphocytes), protein preparations, nucleic acid preparations, etc. A sample may also be obtained by contacting a biological material with an exogenous liquid, resulting in the production of a lavage liquid containing some portion of the contacted biological material. Furthermore, the term "sample" is intended to encompass the primary sample after it has been mixed with one or more additive, such as preservatives, chelators, anti-clotting factors, etc. In

certain embodiments, a fluid sample is a urine sample. In certain embodiments, a preferred solid or semi-solid sample is a stool sample. In certain embodiments, a preferred tissue sample is a biopsy from a tissue known to harbor or suspected of harboring a colon neoplasia. In certain embodiments, a preferred cell culture sample is a sample comprising cultured cells of a colon cancer cell line, such as a cell line cultured from a metastatic colon cancer tumor or a colon-derived cell line lacking a functional TGF-β, TGF-β receptor or TGF-β signaling pathway. A subject is preferably a human subject, but it is expected that the molecular markers disclosed herein, and particularly their homologs from other animals, are of similar utility in other animals. In certain embodiments, it may be possible to detect a marker directly in an organism without obtaining a separate portion of biological material. In such instances, the term sample is intended to encompass that portion of biological material that is contacted with a reagent or device involved in the detection process.

In certain embodiments, a method of the invention comprises detecting the presence of a ColoUp protein in a sample. Optionally, the method involves obtaining a quantitative measure of the ColoUp protein in the sample. In view of this specification, one of skill in the art will recognize a wide range of techniques that may be employed to detect and optionally quantitate the presence of a protein. In preferred embodiments, a ColoUp protein is detected with an antibody. Suitable antibodies are described in a separate section below. In many embodiments, an antibody-based detection assay involves bringing the sample and the antibody into contact so that the antibody has an opportunity to bind to proteins having the corresponding epitope. In many embodiments, an antibody-based detection assay also typically involves a system for detecting the presence of antibody-epitope complexes, thereby achieving a detection of the presence of the proteins having the corresponding epitope. Antibodies may be used in a variety of detection techniques, including enzyme-linked immunosorbent assays (ELISAs), immunoprecipitations, Western blots. Antibody-independent techniques for identifying a protein may also be employed. For example, mass spectroscopy, particularly coupled with liquid chromatography, permits detection and quantification of large numbers of proteins in a sample. Two-dimensional gel electrophoresis may also be used to identify proteins, and may be coupled with mass spectroscopy or other detection techniques, such

as N-terminal protein sequencing. RNA aptamers with specific binding for the protein of interest may also be generated and used as a detection reagent.

In certain preferred embodiments, methods of the invention involve detection of a secreted form of a ColoUp protein or osteopontin, particularly ColoUp1 protein or ColoUp2 protein.

Samples should generally be prepared in a manner that is consistent with the detection system to be employed. For example, a sample to be used in a protein detection system should generally be prepared in the absence of proteases. Likewise, a sample to be used in a nucleic acid detection system should generally be prepared in the absence of nucleases. In many instances, a sample for use in an antibody-based detection system will not be subjected to substantial preparatory steps. For example, urine may be used directly, as may saliva and blood, although blood will, in certain preferred embodiments, be separated into fractions such as plasma and serum.

In certain embodiments, a method of the invention comprises detecting the presence of a ColoUp expressed nucleic acid, such as an mRNA, in a sample. Optionally, the method involves obtaining a quantitative measure of the ColoUp expressed nucleic acid in the sample. In view of this specification, one of skill in the art will recognize a wide range of techniques that may be employed to detect and optionally quantitate the presence of a nucleic acid. Nucleic acid detection systems generally involve preparing a purified nucleic acid fraction of a sample, and subjecting the sample to a direct detection assay or an amplification process followed by a detection assay. Amplification may be achieved, for example, by polymerase chain reaction (PCR), reverse transcriptase (RT) and coupled RT-PCR. Detection of a nucleic acid is generally accomplished by probing the purified nucleic acid fraction with a probe that hybridizes to the nucleic acid of interest, and in many instances detection involves an amplification as well. Northern blots, dot blots, microarrays, quantitative PCR and quantitative RT-PCR are all well known methods for detecting a nucleic acid in a sample.

In certain embodiments, the invention provides nucleic acid probes that bind specifically to a ColoUp nucleic acid. Such probes may be labeled with, for example, a fluorescent moiety, a radionuclide, an enzyme or an affinity tag such as a biotin moiety. For example, the TaqMan® system employs nucleic acid probes that are labeled in such a

way that the fluorescent signal is quenched when the probe is free in solution and bright when the probe is incorporated into a larger nucleic acid.

In certain embodiments, the application provides methods for imaging a colon neoplasia by targeting antibodies to any one of the markers ColoUp1 through ColoUp8 or osetopontin described herein, more preferably the antibodies are targeted to ColoUp3. The markers described herein may be targeted using monoclonal antibodies which may be labeled with radioisotopes for clinical imaging of tumors or with toxic agents to destroy them.

In other embodiments, the application provides methods for administering a imaging agent comprising a targeting moiety and an active moiety. The targeting moiety may be an antibody, Fab, F(Ab)2, a single chain antibody or other binding agent that interacts with an epitope specified by a polypeptide sequence having an amino acid sequence as set forth in SEQ ID Nos: 1-3 and 13-21, preferably an epitope specified by SEQ ID No: 16. The active moiety may be a radioactive agent, such as: radioactive heavy metals such as iron chelates, radioactive chelates of gadolinium or manganese, positron emitters of oxygen, nitrogen, iron, carbon, or gallium, ^{43}K, ^{52}Fe, ^{57}Co, ^{67}Cu, ^{67}Ga, ^{68}Ga, ^{123}I, ^{125}I, ^{131}I, ^{132}I,.or ^{99}Tc. The imaging agent is administered in an amount effective for diagnostic use in a mammal such as a human and the localization and accumulation of the imaging agent is then detected. The localization and accumulation of the imaging agent may be detected by radioscintigraphy, nuclear magnetic resonance imaging, computed tomography or positron emission tomography.

Immunoscintigraphy using monoclonal antibodies directed at the ColoUp markers may be used to detect and/or diagnose colon neoplasia. For example, monoclonal antibodies against the ColoUp marker such as ColoUp3 labeled with .^{99}Technetium, ^{111}Indium, ^{125}Iodine-may be effectively used for such imaging. As will be evident to the skilled artisan, the amount of radioisotope to be administered is dependent upon the radioisotope. Those having ordinary skill in the art can readily formulate the amount of the imaging agent to be administered based upon the specific activity and energy of a given radionuclide used as the active moiety. Typically 0.1-100 millicuries per dose of imaging agent, preferably 1-10 millicuries, most often 2-5 millicuries are administered. Thus, compositions according to the present invention useful as imaging agents

comprising a targeting moiety conjugated to a radioactive moiety comprise 0.1-100 millicuries, in some embodiments preferably 1-10 millicuries, in some embodiments preferably 2-5 millicuries, in some embodiments more preferably 1-5 millicuries.

6. Immunogenic ColoUp proteins

In certain embodiments, the invention relates to methods for identifying ColoUp proteins that elicit an immune response in subjects, such as ColoUp1 through ColoUp8. In one aspect, these immunogenic ColoUp polypeptides have an amino acid sequence that is at least 90%, 95%, or 98-99% identical to the amino acid sequences as set forth in SEQ ID Nos: 1-3 and 13-20. In certain embodiments, such proteins may be suitable as components in a vaccine or for the generation of antibodies that may be used to treat colon cancer.

In certain embodiments, ColoUp proteins that elicit a humoral response may be identified as follows. Sera and/or tissue are obtained from a subject that has been treated for colon cancer by immunotherapy. Proteins from the colon cancer tissue sample will be contacted with antibodies (either purified or in crude serum) to identify proteins that react with the antibodies. The sera or tissue may be obtained, for example, from a center involved in colon cancer immunotherapy.

In one embodiment, ColoUp proteins that elicit a humoral response may be identified by contacting proteins isolated from a colon cancer sample with antibodies obtained from the serum (or simply serum itself or fractions thereof) of a subject having colon cancer. Proteins that react with an antibody from the subject having colon cancer are likely to be proteins that elicit a humoral response. Optionally, the reactivity of proteins is tested against serum or antibodies from a subject not having colon cancer as a comparison, and preferably the antibodies or serum are from the same subject, but at a point in time when the subject did not have colon cancer.

For these methods, proteins may be analyzed in any of the various methods described herein or by other methods that, in view of this specification, are considered to be appropriate by one of skill in the art.

As discussed above, exemplary ColoUp polypeptides include SEQ ID NOs: 1-3 and 15-20. ColoUp polypeptides are further understood to include variants, such as variants of SEQ ID NOs: 1-3 and 15-20.

In another aspect, the invention provides polypeptides that are agonists or antagonists of a ColoUp polypeptide. Variants and fragments of a ColoUp polypeptide may have a hyperactive or constitutive activity, or, alternatively, act to prevent a ColoUp polypeptide from performing one or more functions. For example, a truncated form lacking one or more domain may have a dominant negative effect.

It is also possible to modify the structure of the subject ColoUp polypeptides for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., ex vivo shelf life and resistance to proteolytic degradation in vivo). Such modified polypeptides, when designed to retain at least one activity of the naturally-occurring form of the protein, are considered functional equivalents of the ColoUp polypeptides described in more detail herein. Such modified polypeptides can be produced, for instance, by amino acid substitution, deletion, or addition.

For instance, it is reasonable to expect, for example, that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. conservative mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur -containing = cysteine and methionine. (see, for example,

Biochemistry, 2nd ed., Ed. by L. Stryer, W.H. Freeman and Co., 1981). Whether a change in the amino acid sequence of a polypeptide results in a functional homolog can be readily determined by assessing the ability of the variant polypeptide to produce a response in cells in a fashion similar to the wild-type protein. For instance, such variant forms of a ColoUp polypeptide can be assessed, e.g., for their ability to bind to another polypeptide, e.g., another ColoUp polypeptide. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

This invention further contemplates a method of generating sets of combinatorial mutants of the subject ColoUp polypeptides, as well as truncation mutants, and is especially useful for identifying potential variant sequences (e.g. homologs). The purpose of screening such combinatorial libraries is to generate, for example, ColoUp homologs which can act as either agonists or antagonist, or alternatively, which possess novel activities all together. Combinatorially-derived homologs can be generated which have a selective potency relative to a naturally occurring ColoUp polypeptide. Such proteins, when expressed from recombinant DNA constructs, can be used in gene therapy protocols.

Likewise, mutagenesis can give rise to homologs which have intracellular half-lives dramatically different than the corresponding wild-type protein. For example, the altered protein can be rendered either more stable or less stable to proteolytic degradation or other cellular process which result in destruction of, or otherwise inactivation of the ColoUp polypeptide of interest. Such homologs, and the genes which encode them, can be utilized to alter the levels of a ColoUp protein of interest by modulating the half-life of the protein. For instance, a short half-life can give rise to more transient biological effects and, when part of an inducible expression system, can allow tighter control of recombinant ColoUp polypeptide levels within the cell. As above, such proteins, and particularly their recombinant nucleic acid constructs, can be used in gene therapy protocols.

In similar fashion, homologs of a ColoUp polypeptide can be generated by the present combinatorial approach to act as antagonists, in that they are able to interfere with the ability of the corresponding wild-type protein to function.

Alternatively, other forms of mutagenesis can be utilized to generate a combinatorial library. For example, a ColoUp protein homolog (both agonist and antagonist forms) can be generated and isolated from a library by screening using, for example, alanine scanning mutagenesis and the like (Ruf et al., (1994) Biochemistry 33:1565-1572; Wang et al., (1994) J. Biol. Chem. 269:3095-3099; Balint et al., (1993) Gene 137:109-118; Grodberg et al., (1993) Eur. J. Biochem. 218:597-601; Nagashima et al., (1993) J. Biol. Chem. 268:2888-2892; Lowman et al., (1991) Biochemistry 30:10832-10838; and Cunningham et al., (1989) Science 244:1081-1085), by linker scanning mutagenesis (Gustin et al., (1993) Virology 193:653-660; Brown et al., (1992) Mol. Cell Biol. 12:2644-2652; McKnight et al., (1982) Science 232:316); by saturation mutagenesis (Meyers et al., (1986) Science 232:613); by PCR mutagenesis (Leung et al., (1989) Method Cell Mol Biol 1:11-19); or by random mutagenesis, including chemical mutagenesis, etc. (Miller et al., (1992) A Short Course in Bacterial Genetics, CSHL Press, Cold Spring Harbor, NY; and Greener et al., (1994) Strategies in Mol Biol 7:32-34). Linker scanning mutagenesis, particularly in a combinatorial setting, is an attractive method for identifying truncated (bioactive) forms of a ColoUp polypeptide.

The invention also provides for reduction of the subject ColoUp polypeptides to generate mimetics, e.g. peptide or non-peptide agents, which are able to mimic the behavior or biological activity of the authentic protein. Such mutagenic techniques as described above, as well as the thioredoxin system, are also particularly useful for mapping the determinants of a ColoUp polypeptide which participate in protein-protein interactions involved in, for example, colon cancer.

7. ColoUp nucleic acids

In certain aspects, the invention provides nucleic acids that encode ColoUp proteins. In one aspect, the nucleic acid sequences are at least 90%, 95%, or 98-99% identical to the nucleic acid sequences as set forth in SEQ ID Nos: 4-12. In some embodiments, such nucleic acids include nucleic acids that are differentially expressed in colon cancer samples versus a control sample. In further embodiments, ColoUp nucleic acids encode proteins that are differentially present or absent (or at a different level or in altered form) in the blood of a subject having colon cancer versus a subject not having

colon cancer. In yet additional embodiments, ColoUp nucleic acids include nucleic acids encoding proteins that are differentially expressed (including altered forms etc.) in colon cancer samples versus a control sample. ColoUp nucleic acids are further understood to include nucleic acids that encode variants, such as variants of SEQ ID NOs: 4-12 and nucleic acids encoding SEQ ID NOs: 1-3 and 15-20. Variant nucleotide sequences include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants; and will, therefore, include coding sequences that differ from the nucleotide sequence of the coding sequence due to the degeneracy of the genetic code. In other embodiments, variants will also include sequences that will hybridize under highly stringent conditions to a nucleotide sequence selected from the group consisting of SEQ ID NOs: 4-12 and nucleic acids encoding SEQ ID NOs: 1-3 and 15-20.

One of ordinary skill in the art will understand readily that appropriate stringency conditions which promote DNA hybridization can be varied. For example, one could perform the hybridization at 6.0 x sodium chloride/sodium citrate (SSC) at about 45 °C, followed by a wash of 2.0 x SSC at 50 °C. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50 °C to a high stringency of about 0.2 x SSC at 50 °C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22 °C, to high stringency conditions at about 65 °C. Both temperature and salt may be varied, or temperature or salt concentration may be held constant while the other variable is changed. In one embodiment, the invention provides nucleic acids which hybridize under low stringency conditions of 6 x SSC at room temperature followed by a wash at 2 x SSC at room temperature.

ColoUp nucleic acids include nucleic acids which differ from an identified sequence due to degeneracy in the genetic code. For example, a number of amino acids are designated by more than one triplet. Codons that specify the same amino acid, or synonyms (for example, CAU and CAC are synonyms for histidine) may result in "silent" mutations which do not affect the amino acid sequence of the protein. However, it is expected that DNA sequence polymorphisms that do lead to changes in the amino acid sequences of the subject proteins will exist among mammalian cells. This is particularly likely in the case of nucleic acids derived from cancer samples and proteins that elicit a

humoral response in subjects having colon cancer. One skilled in the art will appreciate that these variations in one or more nucleotides (up to about 3-5% of the nucleotides) of the nucleic acids encoding a particular protein may exist among individuals of a given species due to natural allelic variation. Any and all such nucleotide variations and resulting amino acid polymorphisms are within the scope of this invention.

Another aspect of the invention relates to the use of the isolated nucleic acid in "antisense" therapy. As used herein, antisense therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. binds) under cellular conditions with the cellular mRNA and/or genomic DNA encoding one of the subject ColoUp polypeptides (eg. SEQ ID NOs: 1-3 and 15-20) so as to inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, antisense therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the cellular mRNA which encodes a ColoUp polypeptide. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the mRNA and/or genomic sequences encoding a ColoUp polypeptide. Such oligonucleotide probes are preferably modified oligonucleotide which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and is therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by van der Krol et al., (1988) *Biotechniques* 6:958-976; and Stein et al., (1988) *Cancer Res* 48:2659-2668

Accordingly, the modified oligomers of the invention are useful in therapeutic, diagnostic, and research contexts. In therapeutic applications, the oligomers are utilized in a manner appropriate for antisense therapy in general.

In addition to use in therapy, the oligomers of the invention may be used as diagnostic reagents to detect the presence or absence of the target DNA or RNA sequences to which they specifically bind, such as for determining the level of expression of a gene of the invention or for determining whether a gene of the invention contains a genetic lesion.

8. Identification of candidate colon cancer therapeutics

The present invention also provides assays for identifying therapeutics for treatment of colon cancer. In certain embodiments, such therapeutics may inhibit the expression of a ColoUp protein such as ColoUp 1-8 and osteopontin. Such inhibitory effects can be at the transcriptional level, at the translational level, or at the post-translational level. In certain embodiments, such therapeutics may affect the function of a ColoUp polypeptide such as one selected from the group consisting of SEQ ID NOs: 1-3 and 15-20. For example, such therapeutics may affect the transcriptional factor activity of ColoUp4 and ColoUp5 proteins, or affect the adhesive activity of ColoUp3. In other embodiments, such therapeutics may be targeted to the colon cancer by binding to a ColoUp protein with or without affecting the activity of the ColoUp protein. For example, an aptamer that binds to a ColoUp protein may be conjugated to an anti-cancer therapeutic so as to target the therapeutic to colon cancer cells. In certain embodiments, the anti-ColoUp antibodies as described above may be used in the therapy of colon cancer. Such anti-ColoUp antibodies may be conjugated with radio-nucleotides or cytotoxic agents. Anti-ColoUp antibodies for colon cancer therapy may also include antibodies against cell surface exposed epitopes of a ColoUp protein, for example ColoUp3.

In certain embodiments, candidate therapeutics may be identified on the basis of their ability to modulate the expression of a ColoUp protein. To illustrate, the assay may detect agents which modulate the promoter activity of a ColoUp gene. In certain embodiments, candidate therapeutics may be identified on the basis of their ability to

modulate the binding of a ColoUp polypeptide to an associated protein or ligand. In a further embodiment, the assay detects agents which modulate the intrinsic biological activity of a ColoUp polypeptide. To illustrate, the assay may detect agents which modulate the transcription factor activity of ColoUp4 and ColoUp5 proteins, or the adhesive activity of ColoUp3.

A variety of assay formats will suffice and, in light of the present disclosure, those not expressly described herein will nevertheless be comprehended by one of ordinary skill in the art. Assay formats which approximate such conditions as formation of protein complexes, ligand binding, protein activity, or promoter activity can be generated in many different forms, and include assays based on cell-free systems, e.g. purified proteins or cell lysates, as well as cell-based assays which utilize intact cells. Agents to be tested may be generated in essentially any way, such as, for example, by production in bacteria, yeast or other organisms (e.g. natural products), produced chemically (e.g. small molecules, including peptidomimetics), or produced recombinantly. In a preferred embodiment, the test agent is a small organic molecule, e.g., other than a peptide or oligonucleotide, having a molecular weight of less than about 2,000 daltons.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays of the present invention which are performed in cell-free systems, such as may be developed with purified or semi-purified proteins or with lysates, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the in vitro system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or changes in enzymatic properties of the molecular target.

In an exemplary binding assay, the compound of interest is contacted with a mixture comprising a ColoUp polypeptide and at least one interacting polypeptide or ligand. Detection and quantification of bound ColoUp polypeptide complexes provides a means for determining the compound's efficacy at inhibiting or potentiating interaction.

The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, the binding is quantitated in the absence of the test compound. Complex formation between a ColoUp polypeptide and an interactor may be detected by a variety of techniques, many of which are effectively described above. For instance, modulation in the formation of complexes can be quantitated using, for example, detectably labeled proteins (e.g. radiolabeled, fluorescently labeled, or enzymatically labeled), by immunoassay, or by chromatographic detection. Surface plasmon resonance systems, such as those available from BiaCore, Inc., may also be used to detect protein-protein interaction

Often, it will be desirable to immobilize one of the polypeptides to facilitate separation of complexes from uncomplexed forms of one of the proteins, as well as to accommodate automation of the assay. In an illustrative embodiment, a fusion protein can be provided which adds a domain that permits the protein to be bound to an insoluble matrix. For example, GST-fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with a potential interacting protein, e.g. an 35S-labeled polypeptide, and the test compound and incubated under conditions conducive to complex formation.

ColoUp markers and/or profiles, for example ColoUp3, may be used to screen for therapeutics for colon cancer. Cell surface proteins associated with a disease state may be diminished or eliminated by treatment with certain test compounds. Such test compounds may be useful as therapeutics for the disease state. In addition, certain test compounds may increase the presence of cell surface proteins that are normally present on healthy cells but diminished or absent in diseased cells. Such test compounds may also be useful as therapeutics of colon cancer. Particularly preferred therapeutics will cause the cell surface protein profile of a diseased cell to more closely resemble the cell surface protein profile of a healthy cell.

In further embodiments, the differences between healthy and colon cancer tissue samples may be analyzed to identify targets for therapeutic screening, and a screen may

be designed to identify compounds that bind or otherwise affect the activity of the given target. For example, ColoUp 1-8 proteins and osteopontin are over-expressed in colon cancer. Therapeutics that diminish this over-expression may be useful as colon cancer therapeutics.

In certain embodiments, a method for selecting an appropriate colon cancer therapeutic for a subject is a computer-assisted method. Such a method may comprise obtaining a cell surface protein profile or measuring a marker protein in a sample from a subject. The output signal may then be compared against a database comprising output signal information from a plurality of subjects and further comprising clinical status information from a plurality of subjects. It is contemplated that one may use a computer interface to identify in the database any clinical conditions correlated with the protein profile or marker. Accordingly, one may select a targeted therapeutic to ameliorate or prevent the correlated condition.

9. Tumor vaccines

The treatment of cancer with tumor vaccines has been a goal of physicians and scientists ever since effective immunization against infectious disease with vaccines was developed. In the past, major tumor antigens had not been molecularly characterized. Recent advances are, however, beginning to define potential molecular targets and strategies and this had evolved with the principle that T-cell mediated responses are a useful target for approaches to cancer immunization. In addition, these antigens are not truly foreign and tumor antigens fit more with a self/altered self paradigm, compared to a non-self paradigm for antigens recognized in infectious diseases. Antigens that have been used in the art include the glycolipids and glycoproteins e.g. gangliosides, the developmental antigens, e.g., MAGE, tyrosinase, melan-A and gp75, and mutant oncogene products, e.g., p53, ras, and HER-2/neu. Vaccine possibilities include purified proteins and glycolipids, peptides, cDNA expressed in various vectors, and a range of immune adjuvants.

Any ColoUp protein may be selected for use in a tumor vaccince, although as noted above, ColoUp proteins that elicit a humoral response in subjects having colon cancer are preferred.

Yet another aspect of the present invention relates to the modification of tumor cells, and/or the immune response to tumor cells in a patient by administering a vaccine to enhance the anti-tumor immune response in a host. The present invention provides, for examples, tumor vaccines based on administration of expression vectors encoding a ColoUp gene, or portions thereof, or immunogenic preparations of polypeptides.

In general, it is noted that malignant transformation of cells is commonly associated with phenotypic changes. Such changes can include loss, gain, or alteration in the level of expression of certain proteins. It has been observed that in some situations the immune system may be capable of recognizing a tumor as foreign and, as such, mounting an immune response against the tumor (Kripke, M., Adv. Cancer Res. 34, 69-75 (1981)). This hypothesis is based in part on the existence of phenotypic differences between tumor cells and normal cells, which is supported by the identification of tumor associated antigens (TAAs) (Schreiber, H., et al. Ann. Rev. Immunol. 6, 465-483 (1988)). TAAs are thought to distinguish a transformed cell from its normal counterpart. For example, three genes encoding TAAs expressed in melanoma cells, MAGE-1, MAGE-2 and MAGE-3, have been cloned (van der Bruggen, P., et al. Science 254, 1643-1647 (1991)). That tumor cells under certain circumstances can be recognized as foreign is also supported by the existence of T cells which can recognize and respond to tumor associated antigens presented by MHC molecules. Such TAA-specific T lymphocytes have been demonstrated to be present in the immune repertoire and are capable of recognizing and stimulating an immune response against tumor cells when properly stimulated in vitro (Rosenberg, S.A., et al. Science 233, 1318-1321 (1986); Rosenberg, S.A. and Lotze, M.T. Ann. Rev. Immunol.4, 681-709 (1986)). In the case of melanoma cells both the tyrosinase gene (Brichard, V., et al. J. Exp. Med. 178:489 (1993)) and the Melan-A gene (Coulie et al. J. Exp. Med. 180:35)) have been identified as genes coding for antigens recognized on melanoma cells by autologous cytotoxic lymphocytes.

Induction of T lymphocytes is often a significant early step in a host's immune response. Activation of T cells results in cytokine production, T cell proliferation, and generation of T cell-mediated effector functions. T cell activation requires an antigen-specific signal, often called a primary activation signal, which results from stimulation of a clonally-distributed T cell receptor (TcR) present on the surface of the T cell. This

antigen-specific signal is usually in the form of an antigenic peptide bound either to a major histocompatibility complex (MHC) class I protein or an MHC class II protein present on the surface of an antigen presenting cell (APC). CD4+, helper T cells recognize peptides associated with class II molecules which are found on a limited number of cell types, primarily B cells, monocytes/macrophages and dendritic cells. In most cases class II molecules present peptides derived from proteins taken up from the extracellular environment. In contrast, CD8+, cytotoxic T cells (CTL) recognize peptides associated with class I molecules. Class I molecules are found on almost all cell types and, in most cases, present peptides derived from endogenously synthesized proteins.

The importance of T cells in tumor immunity has several implications which are important in the development of anti-tumor vaccines. Since antigens are processed and presented before they are recognized by T cells, they may be derived from any protein of the tumor cell, whether extracellular or intracellular. In addition, the primary amino acid sequence of the antigen is more important than the three-dimensional structure of the antigen. Tumor vaccine strategies may use the tumor cell itself as a source of antigen, or may be designed to enhance responses against specific gene products. (Pardoll, D. 1993. Annals of the New York Academy of Sciences 690:301).

The present invention provides for various tumor vaccination methods and reagents which can be used to elicit an anti-tumor response against transformed cells which express/display a ColoUp polypeptide, or which have been engineered to present an antigen of a ColoUp polypeptide. In general, the tumor vaccine strategies of the present invention fall into two categories: (1) strategies that use the tumor cell itself as a source of tumor antigen, and (2) antigen-specific vaccine strategies that are designed to generate immune responses against specific antigens of a ColoUp polypeptide.

In general, a ColoUp vaccine polypeptide will include at least a portion of the ColoUp polypeptide, optionally including a site of mutation which, when occurring in the full-length protein, results in loss of its biological activity. Where the colon cancer vaccine comprises a sufficient portion of a ColoUp protein, the protein can be further mutated to render the vaccine polypeptide biologically inactive.

In one embodiment, a tumor cell which otherwise does not express a mutant ColoUp polypeptide can be rendered immunogenic as a target for CTL recognition by

association of a ColoUp vaccine polypeptide. For example, this can be accomplished by the use of gene transfer vectors. Such gene transfer vectors may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively delivering the ColoUp vaccine gene to cells in vivo. Alternatively, cells from the patient or other host organism can be transfected with the tumor vaccine construct ex vivo, allowed to express the ColoUp protein, and, preferably after inactivation by radiation or the like, administered to an individual. In particular, viral vectors represent an attractive method for delivery of tumor vaccine antigens because viral proteins are expressed de novo in infected cells, are degraded within the cytosol, and are transported to the endoplasmic reticulum where the degraded peptide products associate with MHC class I molecules before display on the cell surface (Spooner et al. (1995) Gene Therapy 2:173).

Approaches include insertion of the subject gene into viral vectors including recombinant retroviruses, adenovirus, adeno-associated virus, vaccinia virus, and herpes simplex virus-1, or plasmids. Viral vectors transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramacidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or CaPO4 precipitation carried out in vivo. It will be appreciated that because transduction of appropriate target cells represents the critical first step in gene transfer, choice of the particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g. locally or systemically.

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of a subject ColoUp polypeptide in the tissue of an animal in order to ellicit a cellular immune response. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the vaccine gene by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In another embodiment a mutant ColoUp peptide of the present invention may be directly delivered to the patient. Although such expression constructs as exemplified above have been shown to be an efficient means by which to obtain expression of peptides in the context of class I molecules, vaccination with isolated peptides has also been shown to result in class I expression of the peptides in some cases. For example, the use of synthetic peptide fragments containing CTL epitopes which are presented by class I molecules has been shown to be an effective vaccine against infection with lymphocytic choriomeningitis virus (Schultz et al. 1991. Proc. Natl. Acad. Sci. USA 88:2283) or sendai virus (Kast et al. 1991. Proc Natl Acad Sci. 88:2283). Subcutaneous administration of a CTL epitope has also been found to render mice resistant to challenge with human papillomavirus 16-transformed tumor cells (Feltkamp et al. (1993) Eur. J. Immunol.23:2242-2249). It is contemplated that such peptides may be presented in the context of tumor cell class I antigens or by other, host-derived class I bearing cells (Huang et al. 1994. Science 264:961).

The ColoUp proteins, and portions thereof, may be used in the preparation of vaccines prepared by known techniques (c.f., U.S. Patents 4,565,697; 4,528,217 and 4,575,495). Such polypeptides displaying antigenic regions capable of eliciting protective immune response are selected and incorporated in an appropriate carrier. Alternatively, an antitumor antigenic portion of a ColoUp protein may be incorporated into a larger protein by expression of fused proteins.

The tumor vaccines above may be administered in any conventional manner, including oranasally, subcutaneously, intraperitoneally or intramuscularly. The vaccine may further comprise, as discussed infra, an adjuvant in order to increase the immunogenicity of the vaccine preparation.

In some cases it may be advantageous to couple the ColoUp polypeptide vaccine to a carrier, in particular a macromolecular carrier. The carrier can be a polymer to which the ColoUp polypeptide is bound by hydrophobic non-covalent inneraction, such as a plastic, e.g., polystyrene, or a polymer to which the polypeptide is covalently bound, such as a polysaccharide, or a polypeptide, e.g., bovine serum albumin, ovalbumin or keyhole limpet hemocyanin. The carrier should preferably be non-toxic and non- allergenic. The ColoUp polypeptide may be multivalently coupled to the macromolecular carrier as this

provides an increased immunogenicity of the vaccine preparation. It is also contemplated that the ColoUp polypeptide may be presented in multivalent form by polymerizing the polypeptide with itself.

In addition, the vaccine formulations may also contain one or more stabilizer, exemplary being carbohydrates such as sorbitol, mannitol, starch, sucrose, dextrin, and glucose, proteins such as albumin or casein, and buffers such as alkaline metal phosphate and the like.

The inclusion of CD4+ epitopes in the tumor vaccine in order to further enhance an anti-tumor response is also within the scope of the invention.

In other embodiments, the carcinoma cell itself can be used as the source of antitumor ColoUp antigens. See, for review, Pardoll, D. 1993. Annals of the New York Academy of Sciences 690:301. For example, cells which have been identified through phenotyping as expressing a mutant ColoUp protein can be used to generate a CTL response against a tumor. For example, tumor-infiltrating lymphocytes (TILs) may be derived from tumor biopsies which have such a phenotype. Following such protocols as described by Hom et al. (1991) J Immunotherap 10:153, TILs can be isolated from tumor specimens and grown in the presence of interleukin-2 in order to generate oligoclonal populations of activated T-lymphocytes that are cytolytic to the tumor cells expressing the mutant ColoUp protein.

In other embodiments, whole cell vaccines can be used to treat cancer patients. Such vaccines can include, for example, irradiated autologous or allogenic tumor cells which express (endogenously or recombaintly) a mutant ColoUp polypeptide (or fragment thereof), or lysates of such cells.

In clinical settings, the therapeutic compound of the present invention can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system or peptide can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited

with introduction into the animal being quite localized. For example, the gene delivery vehicle or peptide can be introduced by catheter (see U.S. Patent 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) PNAS 91: 3054-3057). A vaccine gene can be delivered in a gene therapy construct by electroporation using techniques described, for example, by Dev et al. ((1994) Cancer Treat Rev 20:105-115).

The pharmaceutical preparation of the vaccine therapy construct or peptide can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced intact from recombinant cells, e.g. retroviral or adenoviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

Suitable pharmaceutical vehicles for administration to a patient are known to those skilled in the art. For parenteral administration, the ColoUp immunogen will usually be dissolved or suspended in sterile water or saline. For enteral administration, the immunogen will be incorporated into an inert carrier in tablet, liquid, or capsular form. The preparation may also be emulsified or the active ingredient encapsulated in liposome vehicles. The composition or formulation to be administered will, in any event, contain a quantity of the ColoUp polypeptide adequate to achieve the desired immunized state in the subject being treated. The immunogen preparations according to the invention may also contain other peptides or other immunogens.

Suitable carriers may be starches or sugars and include lubricants, flavorings, binders, and other materials of the same nature. For instance, the immunogen can be formulated as a pharmaceutically acceptable acid- or base-addition salt, formed by reaction with inorganic acids such as hydrochloric acid, hydrobromic acid, perchloric acid, nitric acid, thiocyanic acid, sulfuric acid, and phosphoric acid, and organic acids such as formic acid, acetic acid, propionic acid, glycolic acid, lactic acid, pyruvic acid, oxalic acid, malonic acid, succinic acid, maleic acid, and fumaric acid, or by reaction with an inorganic base such as sodium hydroxide, ammonium hydroxide, potassium hydroxide, and organic bases such as mono-, di-, trialkyl and aryl amines and substituted ethanolamines.

The immunogen, which may be coupled to a carrier, is preferably administered after being mixed with immunization adjuvants. Conventional adjuvants include, for example, complete or incomplete Freund's adjuvant, aluminum hydroxide, Quil A, EMA, DDA, TDM-Squalene, lecithin, alum, saponin, and such other adjuvants as are well known to those in the art, and also mixtures thereof. For example, the ColoUp immunogen may be mixed with the N-butyl ester (murabutide) of the muramyl dipeptide (MDP; N-acetyl-glucosamine-3-yl-acetyl-L-alanyl-D-isoglutamine) diluted in a saline solution. The mixture may then be emulsified by means of an equal volume of squalene in the presence of arlacel (excipients). It is also possible to use other adjuvants such as analogues of MDP, bacterial fractions such as streptococcal preparations (OK 432), Biostim (01K2) or modified lipopolysaccharide preparations (LPS), peptidoglycans (N-Opaca) or proteoglycans (K-Pneumonia). In the case of these excipients, water-in-oil emulsions are preferable to oil-in-water emulsions.

In addition to enhancing the immune response against a tumor at its original site, the tumor cell vaccine of the current invention may also be used in a method for preventing or treating metastatic spread of a tumor or preventing or treating recurrence of a tumor. Thus, administration of modified tumor cells or modification of tumor cells in vivo as described herein can provide tumor immunity against cells of the original, unmodified tumor as well as metastases of the original tumor or possible regrowth of the original tumor.

10. Effective Dose

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining The Ld50 (The Dose Lethal To 50% Of The Population) And The Ed50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic induces are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

EXEMPLIFICATION

The invention now being generally described, it will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

Example 1: Selection of eight molecular markers for colon neoplasia

Expression micro-array profiling was used to find genes whose expression was different between normal colon and metastatic colon cancer. Normal colon and metastatic colon cancer samples were analyzed for gene expression using DNA expression microarray techniques that profiled expression patterns of nearly 50,000 genes, ESTs and predicted exons. Analysis of the data identified eight molecular markers for colon neoplasia, as shown in Table 2.

Table 2: Eight Selected Molecular Markers for Colon Neoplasia

Marker Name	Example Sequences (SEQ ID Nos.)	(Median Liver Mets) /(Median Normal Colon)	(Median Liver Mets) /(Median Normal Liver)	(Minimum Liver Mets) /(Maximum Normal Colon)	(Median Met Cell Lines) /(Median Normal Colon)	(Median Met Xenografts) /Median Normal Colon)
ColoUp1	1, 2, 4, 13	13.94	13.94	0.26	14.08	15.48
ColoUp2	3, 5, 14	5.70	5.70	1.00	5.32	1.24
ColoUp3	7, 16	16.36	16.36	0.80	21.50	15.68
ColoUp4	8, 17	4.68	4.68	1.00	4.88	1.56
ColoUp5	9, 18	4.58	4.74	1.15	4.82	4.63
ColoUp6	10, 19	9.52	9.52	0.52	11.58	1.92
ColoUp7	11	9.20	9.20	0.18	4.30	9.00
ColoUp8	12, 20	4.78	4.78	1.27	3.76	2.72

Osteopontin was also identified as a molecular marker having similar characteristics (Example sequences SEQ ID Nos: 6, 15). Each of these molecular markers was subjected to additional analysis in various types of colon neoplasia. In the case of ColoUp1 and ColoUp2, the microarray expression was confirmed by Northern blot and secretion of the protein was established.

Example 2: Expression pattern of ColoUp1 in various cell types.

Shown in Figure 20 is a graphical display of ColoUp1 expression levels measured for different tissue samples. ColoUp1 transcript was essentially undetectable (AI expression levels less than 0) in normal colon epithelial strips (labeled colon epithelial), in normal liver and in colonic muscle (labeled c. muscle). In contrast ColoUp1 expression was clearly detected in premalignant colon adenomas as well as in 90% of Dukes stage B (early node negative colon cancers), Dukes stage C (node positive colon

cancer), Dukes stage D (primary colon cancers with associated metastatic spread) and in colon cancer liver metastasis (labeled liver metastasis). ColoUp1 expression was also demonstrated in colon cancer cell lines (labeled colon cell lines) and in colon cancer xenografts grown in athymic mice (labeled xenografts). The expression in cell lines and xenografts confirms that colon neoplasia cells are the source of ColoUp1 expression in the tumors.

The probe for ColoUp1 was designed to recognize transcripts corresponding to gene KIAA1199, Genbank entry AB033025, Unigene entry Hs.50081. A transcript corresponding to this gene was amplified by RT-PCR from colon cancer cell line Vaco-394. The sequence of this transcript is presented in Figure 3.

Example 3: Confirmed gene expression pattern of ColoUp1

Figure 29 shows a northern analysis using the cloned ColoUp1 cDNA that identifies a transcript running above the large ribosomal subunit (to which the probe cross hybridizes) that is not expressed in normal colon tissue samples and is ubiquitously expressed in a group of colon cancer cell lines.

Figures 29B and 29C show the results of northern analysis of ColoUp1 in normal colon tissue and colon neoplasias from 15 individuals with colon cancers and one individual with a colon adenoma. No normal colon sample expresses ColoUp1. However, expression is see in 13 of 15 colon cancers, and in the one colon adenoma. Expression is seen in cancers arising in both the right and left colon, and in cancers of Dukes Stage B2, C and D.

Example 4: ColoUp1 is a secreted protein

The cloned ColoUp1 colonic transcript was inserted into a cDNA expression vector with a C-terminal T7 epitope tag. Figure 30A shows a summary of the behavior of the tagged protein expressed by transfection of the vector into Vaco400 cells. An anti T7 western blot shows expression of the transfected tagged protein detected in the lysate of a pellet of transfected cells (lane T of cell pellet) which is absent in cells transfected with a control empty expression vector (lane C of cell pellet). Moreover, serial immunoprecipitation and western blotting of T7 tagged protein from media in which

V400 cells were growing (which had been clarified by centrifugation prior to immunoprecipatation) also clearly demonstrates secretion of ColoUp1 protein into the growth medium.

Figure 30B shows the full gels demonstrating expression of tagged 409041 protein in V400 cells demonstrated by western analysis at left and shows detection of secreted 409041 protein in growth media as detected at right by serial immunoprecipitation and western analysis. (Antibody from the high level of serum in which FET cells are grown blocked the ability of staphA conjugated beads to precipitate anti-T7 bound to 409041 in growth media from FET cells).

Example 5: Expression pattern of ColoUp2 in various cell types.

Shown in Figure 21 is the graphical display of ColoUp2 expression levels measured for different samples analyzed. ColoUp2 transcript was essentially undetectable (AI expression levels less than 0) in normal colon epithelial strips (labeled colon epithelial), in normal liver and in colonic muscle (labeled c. muscle). In contrast ColoUp2 expression was clearly detected in premalignant colon adenomas as well as in 90% of Dukes stage B (early node negative colon cancers), Dukes stage C (node positive colon cancer), Dukes stage D (primary colon cancers with associated metastatic spread) and in colon cancer liver metastasis (labeled liver metastasis). ColoUp2 expression was also demonstrated in colon cancer cell lines (labeled colon cell lines) and in colon cancer xenografts grown in athymic mice (labeled xenografts). The expression in cell lines and xenografts confirms that colon neoplasia cells are the source of ColoUp2 expression in the tumors.

Probe ColoUp2 was designed to recognize transcripts corresponding to a noncoding EST, Genbank entry AI357412, Unigene entry Hs.157601. By 5' RACE, database assembly, and ultimately RT-PCR, we cloned from a colon cancer cell line a novel protein encoding RNA transcript whose noncoding 3' UTR was shown to correspond to the ColoUp2 specified EST. This full length coding sequence was determined by RT-PCR amplification from colon cancer cell line Vaco503 and sequences are provided in Figure 4.

ColoUp2 is a "class identifier" (that is, it is higher in all colon cancer samples than in all normal colon samples), it is not-expressed in normal body tissues and it contains a signal sequence predicting that the protein product will be secreted (as well as several other recognizable protein motifs including domains from the epidermal growth factor protein and from the Von Willebrands protein).

Example 6: Confirmed gene expression pattern of ColoUp2

Figure 31 shows a northern analysis using the cloned ColoUp2 cDNA that identifies a transcript running above the large ribosomal subunit (to which the probe cross hybridizes) that is not expressed in normal colon tissue samples and is expressed in the majority of group of colon cancer cell lines. Panel A of the figure shows the northern hybridization. The red arrow designates the ColoUp2 transcript. Above each lane is the name of the sample and the level (in parenthesis) of ColoUp2 expression recorded. The black arrow designates the cross hybridizing ribosomal large subunit. Panel B shows the eithidum bromide stained gel corresponding to the blot, and the black arrows designate the large and small ribosomal subunits.

Example 7: ColoUp2 is a secreted protein

The cloned ColoUp2 colonic transcript was inserted into a cDNA expression vector with a C-terminal V5 epitope tag. Figure 32 shows a summary of the behavior of the tagged protein expressed by transfection of the vector into SW480 and Vaco400 cells. An anti V5 western blot shows (red arrows) expression of the transfected tagged protein detected in the lysate of a pellet of transfected cells (lysates western panel, lanes labeled ColoUp2/V5) which is absent in cells transfected with a control empty expression vector (lanes labeled pcDNA3.1). Moreover, serial immunoprecipitation and western blotting of V5 tagged protein from media in which V400 and SW480 cells were growing (which had been clarified by centrifugation prior to immunoprecipatation) also clearly demonstrates secretion of the ColoUp2 protein into the growth medium (panel labeled medium IP-western). Antibody bands from the immunoprecipitation are also present on the IP-western blot. Detection of secreted ColoUp2 protein was shown in cells assayed both 24 hours and 48 hours after transfection.

Example 8: Expression pattern of ColoUp3 – ColoUp8 and osteopontin in various cell types.

Shown in Figures 22-28 are the graphical displays of ColoUp3 – ColoUp8 and osteopontin expression levels measured for different samples analyzed.

Example 9: Confirmed gene expression pattern of ColoUp5

Shown in Figure 33 is a northern blot showing that ColoUp5 is expressed in colon cancer cell lines and not expressed in non-neoplastic material. Figure 33 shows two northern blot analysis of ColoUp5 mRNA levels in normal colon tissues and a group of colon cancer cell lines (top panels). The bottom panels show the ethidium bromide stained gel corresponding to the blot. Homologs for ColoUp5 are found in other mammals, including mouse and rat, and sequence alignments are shown in Figures 34 and 35.

Example 10: Detection of xenograft derived ColoUp1 and ColoUp2 proteins circulating in the blood of mice.

To determine that ColoUp1 and ColoUp2 proteins are effective serologic markers of colon neoplasia, we derived transfected cell lines that stably expressed and secreted V5-epitope tagged ColoUp1 and ColoUp2 proteins. These cells lines were then injected into athymic mice and grown as tumor xenografts. Mice were sacrificed and serum was obtained. V5 tagged proteins were then precipitated from the serum using beads conjugated to anti-V5 antibodies. Precipitated serum proteins were run out on SDS-PAGE, and visualized by western blotting using HRP-conjugated anti-V5 antibodies (thereby eliminating visualization of any contaminating mouse immunoglobulin). Figure 36 shows detection of circulating ColoUp2 protein in mouse serum. The ColoUp2 protein is secreted as 2 bands of 85KD and 55KD in size, of which the 55KD band predominates in the serum. The 55KD band is presumably a processed form of the 85KD band. This observation demonstrates that, in this mouse model, ColoUp2 is indeed a secreted marker of colon cancers and adenomas, and that ColoUp2 can gain access to and circulate stably in patient serum. This observation provides the surprising result that a

processed fragment of ColoUp2 is the predominant serum form of the protein and therefore detection reagents targeted to this portion would be particularly suitable for diagnostic testing.

A time course experiment showed that ColoUp2 protein was detectable in mouse blood at the earliest time assayed, 1 week after injection of ColoUp2 secreting colon cancer cells, at which time xenograft tumor volume as only $100mm^3$.

Similar observations were also made for ColoUp1, as shown in Figure 37.

Example 11: Purification of ColoUp1 and ColoUp2 proteins.

In order to develop monoclonal antibodies against native ColoUp1 and ColoUp2 proteins, we devised a protocol for purification on Ni-NTA agarose (QIAGEN) nickel beads of recombinant His tagged ColoUp1 and ColoUp2 proteins from the media supernate of SW480 cells engineered to express these proteins. Currently we have purified both ColoUp1 and ColoUp2 proteins to sufficient purity to generate antibodies. As shown in Figure 38, a Coomassie blue stained gel of purified ColoUp2 shows only the 85KD and 55KD size bands that correspond to the tagged ColoUp2 proteins visualized on western blot. Similarly, a Coomassie blue stained gel of purified ColoUp1 shows the preparation is highly purified and composed of a single 180KD band that corresponds perfectly to the size band seen on western blotting of the epitope tagged ColoUp1 protein. Thus we have purified ColoUp2 and ColoUp1 to sufficient homogeneity and yield. Scaled up purification of these proteins from a 50 liter media preparation should yield 2.5 mg of protein, more than adequate for immunizing mice and screening fusion supernates for development of monoclonal antibodies specific for native ColoUp1 and ColoUp2.

Example 12: Measuring apical and basolateral secretion of ColoUp1 and ColoUp2.

We expected that ColoUp2 will serve as a serologic marker detection not only of colon cancers but also of large colon adenomas that also express ColoUp2. Adenomas, unlike colon cancers, are non-invasive. Thus, for adenomas to move ColoUp2 proteins into the circulation they would need to secrete this protein from the basolateral cell surface facing capillaries and lymphatics, rather than from the apical cell surface facing the colon lumen. To determine the polarity of ColoUp2 secretion we transiently

transfected a monolayer of polarized Caco2 colon cancer cells with an expression vector for V5-epitope tagged ColoUp2 protein. This cell monolayer was grown in transwell dishes on filters that separate an upper transwell chamber (representing media exposed to the apical surface of the monlayer) from a lower transwell chamber (representing media exposed to the basolateral surface of the monolayer). Integrity of the sealing of the monolayer was assayed by measuring electrical resistance across the filters, and efficiency of transient transfection was monitored by expression of a gfp marker. Media from upper and lower chambers was harvested at 24, 48, 72, and 96 hours post transfection, and secreted tagged ColoUp2 protein was detected by western analysis directed against the V5 epitope tag. As Figure 39 shows, characteristic 85KD and 55KD secreted forms of ColoUp2 were detected in media sampling the basolateral monolayer compartment at all time points assayed. At a single time point, 48 hours, ColoUp2 was additionally detected in media representing the apical secretion face; however, a dip in the transfilter electrical resistance at 48 hours suggests the likelihood of some leaking across the monolayer at this time point. Certainly, the data clearly shows secretion of ColoUp2 into the basolateral monolayer compartment, and hence establishes ColoUp2 as demonstrating the requisite biology for a candidate serologic marker of colon adenomas.

As was done for ColoUp2, ColoUp1 expression vectors were used to transiently transfect Caco2 cell monolayers grown on transwell filters. Secretion of ColoUp1 was then assayed in media collected respectively from the upper and lower transwell chambers. Western blot assays demonstrated equal secretion of ColoU1 from both apical and basolateral monolayer surfaces. Studies of ColoUp1 were done in parallel with those of ColoUp2, and electrical resistance of the ColoUp1 monolayers exceeded that of the ColoUp2 monolayers, supporting that the ColoUp1 transfected monolayers were well sealed. Additionally, levels of secreted ColoUp1 protein were similar to those of secreted ColoUp2, suggesting that ColoUp1 secretion by both apical and basolateral compartments was not simply due to overexpression.. Accordingly, we predict that native ColoUp1 protein is likely secreted at least in part from the basolateral epithelial face, and hence should be detectable as a serologic marker of large colon adenomas.

Example 13: Determining the sequence of the 55 kDa ColoUp2 fragment

The protein sequence of C-terminal fragment of ColoUp2 that is secreted by human cell lines and detected as predominant fragment in blood (488 aa) was determined. As described above, we have found on western blots and on purified preparations of C-terminal epitope tagged (V5-His epitope) ColoUp2 protein secreted by transfected human colon cancer cells, both a full sized band of approximately 90 kDa and a smaller approximately 55 kDa C-terminal fragment (as demonstrated by the retention of the C-terminal epitope tag). Moreover, when these cells were injected into athymic mice, the 55 kDa C-terminal tagged protein was the predominant species detected as circulating in the mouse blood, when mouse serum is analyzed by serial immunoprecipitation and western blot analysis directed against the V5 tag. The precise location of the cleavage site accounting for the C-terminal fragment was established by excising the acrylamide gel band containing the purified C-terminal fragment and performing mass spectroscopy analysis of tryptic fragments from the protein. A peptide of sequence AVLAAHCPFYSWK was present only in the digest of the 55KD fragment, but was absent from the digest of the full length protein, demonstrating that this peptide corresponded to the unique amino terminus of the 55KD fragment. The complete sequence of the 55KD C-terminal fragment is shown in Figure 41.

INCORPORATION BY REFERENCE

All publications and patents mentioned herein are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

EQUIVALENTS

While specific embodiments of the subject invention have been discussed, the above specification is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification and the claims below. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.